

# Gjenfortelling som redskap i kartleggingen av førskolebarns språkferdigheter

*En undersøkelse av fireåringers prestasjoner på den narrative testen Bus Story, med særlig henblikk på sammenhenger med ordforråd og grammatiske ferdigheter*

**Stig Berdal**



Masteroppgave ved Det utdanningsvitenskapelige Fakultet, Institutt  
for Spesialpedagogikk

**UNIVERSITETET I OSLO**

01.11.2009

---

### Sammendrag:

Oppgaven er skrevet i tilknytning til det longitudinelle forskningsprosjektet Child Language and Learning ved Institutt for Spesialpedagogikk (ISP), Universitetet i Oslo. Prosjektets formål er å frambringe kunnskap om barns språkutvikling. Ved hjelp av et omfattende testbatteri innhentes informasjon om en rekke språklig-kognitive ferdigheter hos et representativt utvalg norske barn mellom fire og åtte år. Studenter ved ISP har deltatt i testingen, og fått anledning til å bruke innsamlede data i egne masteroppgaver. Den narrative testen Bus Story inngikk i testbatteriet som ble brukt da barna var fire. Barna skal i denne testen gjenfortelle en kort historie, med støtte i en bildeserie som viser historiens handling. Det beregnes to skårer, en for hvor godt barna gjengir historiens innhold, og en for lengden i setningene de bruker. Denne masteroppgaven tar opp spørsmålet om hvordan disse skårene kan tolkes. Hva slags informasjon ligger i skårene, hvilke slutninger gir de grunnlag for å trekke med hensyn til barnas språkferdigheter? Dermed berøres spørsmålet om hvordan denne testen kan brukes i en kartlegging av fireåringers språk. To problemstillinger er formulert. Den første gjelder hvorvidt skårer på Bus Story kan antas å være en god *indikator* på språklige ferdigheter hos fireåringers, den andre gjelder i hvilken grad skårene kan antas å *predikere* fireåringers språkutvikling.

**Metode:** Oppgaven har et kvantitativt, ikke-eksperimentelt design. Problemstillingens to spørsmål søkes besvart gjennom en undersøkelse av empiriske sammenhenger mellom barnas prestasjoner på Bus Story, og deres prestasjoner på tester som kartlegger ordforråd og grammatiske ferdigheter, både samtidig og ett år senere. Det legges til grunn at sammenhengene bør være av en viss styrke for at Bus Story skal kunne sies å indikere språkferdigheter og predikere språkutvikling på en god måte. For å måle ordforråd og grammatiske ferdigheter er anerkjente og relevante tester benyttet. Sammenhengen mellom Bus Story-resultater og generelt evnenivå samt umiddelbar hukommelse for talt språk er også undersøkt, da også dette har betydning for hvordan testresultatene kan tolkes.

**Hovedresultater:** De empiriske undersøkelsene viser svake sammenhenger mellom fireåringenes resultater på Bus Story og deres ordforråd og grammatiske ferdigheter.

---

Den sterkeste korrelasjonen finner vi mellom informasjonsskåren på Bus Story og Bildebenevning, testen for ekspressivt ordforråd, hvor  $r=.374$ ,  $p<.01$ .

Regresjonsanalyser viser at Bildebenevning forklarer 14% av variansen i barnas prestasjoner på Bus Story når den innføres først i analysen. De øvrige språktestene bidrar da ikke med ytterligere forklart varians. Det gjenstår med andre ord 86 % uforklart varians når språktestene har forklart det de kan. Også sammenhengene mellom Bus Story og testene for generelt evnenivå og hukommelse for talt språk, hhv. Terningmønster og Setningsminne, er svake, med korrelasjonskoeffisienter ( $r$ ) på henholdsvis  $.178$ ,  $p<.05$  og  $.297$ ,  $p<.01$ . Når disse testene innføres i en regresjonsanalyse etter Bildebenevning, forklares kun ytterligere 3.7% varians. Det er likeledes svake sammenhenger mellom fireåringenes prestasjoner på Bus Story, og deres resultater på tester for ordforråd og grammatikk ett år senere.

Takk:

En stor takk til professor Sol Lyster for sjenerøs og konstruktiv veiledning. Takk til medstudent Jorun for god hjelp i arbeidets siste fase, og til Christiane for mange gode samtaler underveis og godt følge gjennom masterstudiet.

## Oversikt over tabeller og figurer

Figur 2-1: Toulmins modell. ....	22
Tabell 3-1: Alfaverdier T1 .....	45
Tabell 3-2: Alfaverdier T2 .....	45
Figur 4-1: Histogram som viser fordelingen på Bus Story .....	55
Tabell 4-1: Korrelasjoner mellom Bus Story informasjon og andre språktester administrert ved fire års alder. ....	56
Figur 4-2 - Spredningsdiagram for sammenhengen mellom Bus Story og Bildebenevning. ....	57
Figur 4-3 – Spredningsdiagram for sammenhengen mellom Bus Story og Grammatic Closure. ....	58
Figur 4-4 – Spredningsdiagram for sammenhengen mellom Bus Story Informasjon og BPVS.....	58
Figur 4-5: Spredningsdiagram for sammenhengen mellom Bus Story og Trog-2.....	59
Tabell 4-2: Korrelasjoner mellom Bus Story og Terningmønster og Setningsminne	60
Tabell 4-3: Hierarkiske regresjonsanalyser med Bus Story som avhengig variabel og Bildebenevning, Grammatic Closure, BPVS, Trog 2-items som uavhengige variabler. ....	61
Tabell 4-4: Hierarkiske regresjonsanalyser med Bus Story som avhengig variabel og Setningsminne, Terningmønster og Bildebenevning som uavhengige variabler.....	63
Tabell 4-5: Korrelasjoner mellom Bus Story og tester for ordforråd og grammatikk administrert ett år senere. ....	65
Tabell 4-6: Gruppen som skåret null poeng på Bus Story Informasjon (0-gruppen) sammenlignet med hele utvalget mhp resultater på språktester, setningsminne og terningmønster (alder fire år) .....	66
Tabell 4-7: Gruppen som skåret 0 poeng på Bus Story ved fire års alder (0-gruppen) sammenlignet med hele utvalget mhp resultater på språktester ett år senere .....	68

# Innhold

<b>INNHold .....</b>	<b>5</b>
<b>1. INNLEDNING: BAKGRUNN OG FORMÅL, PROBLEMSTILLING OG OPPGAVENS OPPBYGNING .....</b>	<b>7</b>
1.1 BAKGRUNN OG FORMÅL .....	7
1.2 PROBLEMSTILLING .....	9
1.3 OPPGAVENS OPPBYGNING .....	11
<b>2. TEORETISK OG EMPIRISK RAMME .....</b>	<b>13</b>
2.1 VALIDITET OG TESTBRUK .....	13
2.2 INNHOLDSVALIDITET OG BEGREPSVALIDITET .....	17
2.3 STRUKTUREN I ET VALIDITETSARGUMENT – TOULMINS MODELL .....	20
2.3.1 Vurdering av Bus Story-testens validitet i lys av Toulmins modell .....	22
2.3.2 Testskårenes validitet som uttrykk for norske fireåringers språklige ferdigheter .....	24
2.3.3 Testskårenes validitet som prediktorer av norske fireåringers språkutvikling .....	29
2.4 FORELØPIG OPPSUMMERING .....	33
<b>3. METODE .....</b>	<b>34</b>
3.1 DESIGN .....	34
3.2 UTVALG .....	36
3.3 TESTER BRUKT I DENNE UNDERSØKELSEN .....	38
3.3.1 Bus Story .....	38
3.3.2 Bildebenevning (WPPSI-III+EVT 2) .....	38
3.3.3 Grammatic Closure .....	39
3.3.4 British Picture Vocabulary Scale II (BPVS) .....	39

---

3.3.5	<i>Test for Reception of Grammar-2 (TROG-2)</i> .....	39
3.3.6	<i>Ordforråd (WPPSI-R)</i> .....	40
3.3.7	<i>Terningmønster</i> .....	41
3.3.8	<i>Setningsminne</i> .....	41
3.4	UNDERSØKELSENS VALIDITET OG RELIABILITET .....	41
3.4.1	<i>Reliabilitet</i> .....	42
3.4.2	<i>Validitet</i> .....	47
3.5	ETIKK.....	52
<b>4.</b>	<b>RESULTATER</b> .....	<b>54</b>
4.1	FORHOLDET MELLOM INFORMASJONSSKÅREN OG SETNINGSLENGDESKÅREN .....	54
4.2	FIREÅRINGENES RESULTATER PÅ BUS STORY .....	55
4.3	SAMMENHENGER MELLOM BUS STORY OG ANDRE SPRÅKTESTER .....	56
4.4	SAMMENHENGER MELLOM BUS STORY OG KOGNITIVE TESTER.....	59
4.5	SPRÅKLIGE VARIABLERES BIDRAG TIL Å FORKLARE VARIASJONEN I RESULTATER PÅ BUS STORY	60
4.6	KOGNITIVE VARIABLERES BIDRAG TIL Å FORKLARE VARIASJONEN I RESULTATER PÅ BUS STORY	62
4.7	BUS STORY-TESTENS PREDIKTIVE VALIDITET: SAMMENHENGER MELLOM BUS STORY OG SPRÅKTESTER ADMINISTRERT ETT ÅR SENERE .....	64
4.8	KJENNETEGN VED GRUPPEN AV BARN SOM SKÅRET 0 POENG PÅ BUS STORY .....	65
<b>5.</b>	<b>OPPSUMMERING AV FUNN, DISKUSJON OG AVSLUTNING</b> .....	<b>69</b>
5.1	OPPSUMMERING AV FUNN .....	69
5.2	DISKUSJON .....	70
5.3	AVSLUTNING: IMPLIKASJONER OG FREMTIDIGE FORSKNINGSSPØRSMÅL.....	76
	<i>Kildeliste</i> .....	78

# 1. Innledning: Bakgrunn og formål, problemstilling og oppgavens oppbygning

## 1.1 Bakgrunn og formål

Oppgaven er skrevet i tilknytning til det longitudinelle forskningsprosjektet *Child Language and Learning* (heretter kalt Språkprosjektet) ved Institutt for Spesialpedagogikk (ISP), Universitetet i Oslo. Formålet med Språkprosjektet er å frambringe kunnskap om språkutviklingen hos barn mellom fire og åtte år. Språkutviklingen hos barn med minoritetsbakgrunn, Down syndrom, Cochlea-implantat og spesifikke språkvansker skal studeres opp mot en normgruppe på cirka 200 barn med normal språkutvikling. Ved hjelp av et omfattende testbatteri innhentes informasjon om et bredt spekter av språklig-kognitive ferdigheter, slik som ordforråd, grammatiske ferdigheter, verbale minnefunksjoner og fonologisk bevissthet. Studenter ved ISP deltar i testingen av barna i normgruppen, og får bruke innsamlede data i egne masteroppgaver. Denne oppgaven er skrevet etter andre testrunde, altså når barna er omkring fem år. Det blir imidlertid benyttet data fra begge testrunder.

Mens Språkprosjektets overordnede mål er å øke kunnskapen om barns språk og språkutvikling, fokuseres det i denne oppgaven på *hvordan* man kan framskaffe slik kunnskap, nærmere bestemt hvilken rolle språktester kan spille i denne sammenhengen. Oppgavens formål er å bidra med prinsipiell refleksjon omkring bruk av tester og tolking av testresultater, og å drøfte hvilken informasjon en bestemt test kan tenkes å gi. Testen det dreier seg om er den narrative testen Bus Story (Renfrew, 1997). Den administreres ved at testlederen forteller barnet en kort historie, samtidig som barnet får se en serie på tolv bilder som illustrerer historiens handling. Deretter skal barnet gjenfortelle historien, med bildene som støtte. Barnets gjenfortelling skåres på to måter: Det gis en skåre for hvor fullstendig barnet formidler historiens innhold, og en skåre for lengden i setningene som brukes.<sup>1</sup> Det sentrale spørsmålet i

---

<sup>1</sup> En fylldigere presentasjon av testen gis senere i oppgaven, under punkt 2.4.

denne oppgaven er hvordan disse skårene kan tolkes. Bus Story er tatt inn i Språkprosjektet fordi den er en av de mest brukte og anerkjente narrative testene, både klinisk og i forskningssammenheng (Howlin & Kendall, 1991; Botting, 2002). Mange forskere anbefaler narrative tester som redskap i kartleggingen av barns språk og språkutvikling (Bishop & Edmundson, 1987; Howlin & Kendall, 1991; Paul & Smith, 1993; Botting, 2002; Fey, Catts, Proctor-Williams, Tomblin & Zhang, 2004.) Det legges til grunn at det å fortelle eller gjenfortelle en historie er en kompleks språklig-kognitiv oppgave, som tapper et bredt spekter av ferdigheter, blant annet ordforråd og grammatisk kompetanse (Paul & Smith, 1994; Renfrew, 1997; Botting, 2002). Dermed kan man ved hjelp av en enkelt test få mye og viktig informasjon. Bus Story presenteres i manualen som en screeningtest, som gjør det mulig på en rask og enkel måte å innhente informasjon om viktige sider ved barns språkferdigheter, i første rekke ekspressive, men også reseptive (Renfrew, 1997). Narrative testers prediktive nytte har vært gjenstand for særskilt interesse. I to mye siterte undersøkelser (Bishop & Edmundson, 1987; Stothard m.fl., 1998) viste det seg at Bus Story hadde god prediktiv validitet når det gjaldt den videre språkutviklingen til barn med spesifikke språkvansker. Mer generelt pekes det i noen undersøkelser på mulige sammenhenger mellom førskolebarns narrative ferdigheter og deres senere kognitive og språklige utvikling, og deres skolefaglige mestring overhodet (Feagan & Applebaum, 1986; Liles, 1993; Paul & Smith, 1994; Botting 2002). Samtidig som narrative testers potensielle nytteverdi ofte framheves, påpeker flere forskere at det kan være problematisk å spesifisere hva det egentlig er disse testene måler (Pankratz m. fl., 2007). Blant annet fordi så mange delferdigheter er involvert i narrativ produksjon, kan det være vanskelig å tolke resultatet på en narrativ test, hva enten dette anses som svakt eller sterkt. Hva er det som har innvirket på resultatet? Det kan med andre ord bli vanskelig å svare presist på det i kartleggingssammenheng helt sentrale spørsmålet: Hva slags kunnskap får vi egentlig ved å bruke Bus Story? Eller mer spesifikt: Hva sier testresultatet egentlig om barnas språkmestring? De dataene som er innsamlet gjennom Språkprosjektet, gjør det mulig å nærme seg dette spørsmålet empirisk. Det kan naturligvis ikke gis noe endelig eller uttømmende svar, men det er mulig å undersøke noen empiriske sammenhenger som kan bidra til å



kaste lys over spørsmålet. Det kan gjøres ved å sammenligne barnas prestasjoner på Bus Story med deres språkferdigheter for øvrig, slik disse lar seg måle ved hjelp av noen av de andre testene som inngår i testbatteriet. Er det for eksempel en tendens til at de barna som gjør det bra på Bus Story, også har et godt ordforråd og gode grammatiske ferdigheter, og hvor sterk er i så fall denne tendensen?

## 1.2 Problemstilling

Tema for oppgaven er hva barns prestasjoner på Bus Story kan si om deres språklige ferdigheter. I denne undersøkelsen benyttes data fra testing av et representativt utvalg norske fireåringer med normal språkutvikling. Oppgaven har to problemstillinger. Den første lyder:

**Er skårene på Bus Story gode indikatorer på språklige ferdigheter hos norske fireåringer?**

Dette spørsmålet forsøkes besvart ved å undersøke sammenhenger mellom barnas prestasjoner på Bus Story og deres ordforråd og grammatiske ferdigheter. Dersom sammenhengene er sterke, kan det sies at Bus Story gir en god indikasjon på barnas språklige ferdigheter. Testresultatene synes i så fall å indikere tilstedeværelsen av ferdigheter som etter allmenn oppfatning utgjør kjerneelementer i språklig kompetanse (Bloom & Lahey, 1978). Dersom sammenhengene derimot er svake, blir en slik konklusjon naturligvis problematisk. Samtidig aktualiseres da spørsmålet om hva testen faktisk måler, og hvilke faktorer som innvirker på testresultatet. Uansett hvor sterke eller svake sammenhengene viser seg å være, vil funnene ha relevans for hvilke slutninger som kan trekkes på grunnlag av skårene på Bus Story. Som mål for ordforråd og grammatikk brukes skårer på anerkjente og mye brukte tester. Disse beskrives nærmere i oppgavens metodekapittel. Oppgaven vil også ta opp spørsmålet om i hvor stor grad testresultatene gjenspeiler kognitive ferdigheter, som generelt evnenivå og umiddelbar hukommelse for talt språk, da også dette har relevans for hvordan skårene kan tolkes. Empiriske sammenhenger mellom barnas resultater på

---

Bus Story og deres skårer på tester som måler slike ferdigheter blir undersøkt og diskutert. Den andre problemstillingen lyder:

**Hvor godt kan resultater på Bus Story predikere språklig utvikling for norske fireåringer?**

For å få et svar på dette spørsmålet, undersøkes korrelasjoner mellom fireåringenes skårer på Bus Story, og deres skårer på et utvalg tester av ordforråd og grammatikk ett år senere. Korrelasjonenes styrke sier noe om i hvilken grad Bus Story kan predikere fireåringenes språkutvikling.

Den empiriske undersøkelsen bygger på en relativt fyldig teoridel, hvor de to problemstillingene blir drøftet rent teoretisk. Det innebærer at en viktig del av oppgavens analysearbeid gjøres i den innledende teoridelen. Et viktig prinsipielt spørsmål som berøres i oppgaven, er nettopp forholdet mellom en teoretisk og en empirisk vurdering av hva slags informasjon som kan antas å ligge i en testskåre. Dette forholdet undersøkes i denne oppgaven med henblikk på Bus Story: Hva slags informasjon kan testskårene på teoretisk grunnlag antas å gi, og hva viser en empirisk undersøkelse?

Det er viktig å presisere at det er *skårene* på Bus Story som står i fokus i denne oppgaven, altså hvilken informasjon fireåringenes skårer kan antas å være bærere av. Det er med andre ord ikke en generell vurdering av testens kliniske brukbarhet som tilstrebes. Det kan være mange grunner til at det er nyttig å bruke gjenfortelling som et verktøy i språkkartlegging, ikke minst i en dynamisk preget kartleggingsprosess. Barnets testrespons utgjør en forholdsvis fyldig språkprøve, og kan sette testleder på sporet av forhold som bør undersøkes nærmere. Responsen kan for eksempel inneholde påfallende trekk av fonologisk eller grammatisk art. Den narrative formen kan også gi testleder god anledning til å observere barnets lærings- og problemløsningsstrategier, for eksempel hvordan han eller hun nyttiggjør seg støtte og hjelp. Da vil det imidlertid være tale om en kvalitativ tolkning av barnets atferd i

testsituasjonen og av resultatene. Diskusjonen i denne oppgaven gjelder hvilke slutninger som kan trekkes på grunnlag av skårene for innhold og setningslengde, og hvordan man kan begrunne disse slutningene. Det dreier seg om det som i den engelskspråklige litteraturen kalles *score-based inferences* (Messick, 1980).

### 1.3 Oppgavens oppbygning

Som bakgrunn og ramme for den empiriske undersøkelsen diskuteres begrepet validitet. Validitet betyr gyldighet, og betegner i denne sammenhengen gyldigheten i de slutningene man trekker på grunnlag av et testresultat. Et hovedspørsmål når det gjelder testvaliditet er under hvilke forutsetninger man med rimelig grad av sikkerhet kan si hva et testresultat innebærer. Hva skal til for at et utsagn av typen: ”Testresultatene viser at...” har overbevisningskraft? Kan det argumenteres tilstrekkelig overbevisende rent teoretisk for at en test gir en bestemt type informasjon? Eller må en slik påstand underbygges empirisk? Disse spørsmålene knytter an til distinksjonen mellom innholdsvaliditet og begrepsvaliditet (Messick, 1980; Plante & Vance, 1994; Kleven, 2002a). Begge termene betegner i hvilken grad en test gir informasjon om det innholdsområdet den er ment å si noe om, altså om testresultatet er et gyldig uttrykk for hva testpersonen vet eller kan innenfor det aktuelle området. Vurdering av innholdsvaliditet baserer seg på en skjønnsmessig, teoretisk fundert oppfatning av hva testen måler, mens vurdering av begrepsvaliditet i tillegg støtter seg på empiriske undersøkelser: De sammenhengene man antar gjør seg gjeldende mellom testresultatet og visse ferdigheter eller egenskaper hos testpersonen, søkes påvist empirisk (Kleven, 2002a). I praksis betyr det at resultater fra testen vises å stå i en bestemt relasjon til resultater fra andre, relevante målinger. Som neste ledd i oppgavens teoretiske del presenteres en modell som kan brukes for å vurdere holdbarheten i en påstand om hva et gitt testresultat innebærer (Toulmin, 2003). Hensikten med å bruke modellen er å synliggjøre strukturen og innholdet i det resonnementet som en slik påstand bygger på. Modellen blir så brukt i en vurdering av om de skårene som norske fireåringer oppnår på Bus Story, kan ses som 1) gode

indikatorer på deres ordforråd og grammatiske ferdigheter, og 2) gode prediktorer av deres språklige utvikling. Vurderingens konklusjon er at det ikke kan argumenteres tilstrekkelig overbevisende på rent teoretisk grunnlag for at testskårene kan antas å gi denne type informasjon. De samme to forholdene blir så vurdert empirisk. Det skjer gjennom en undersøkelse av statistiske sammenhenger mellom resultater på Bus Story og resultater på et utvalg andre tester. Oppgaven avsluttes med en oppsummering og drøfting av de empiriske resultatene, og implikasjoner for bruken av Bus Story blir diskutert. Endelig blir mulige fremtidige forskningsspørsmål antydnet.

## 2. Teoretisk og empirisk ramme

### 2.1 Validitet og testbruk

Generelt sett er hovedformålet med testing av barn i pedagogisk sammenheng å bidra til optimale lærings- og utviklingsbetingelser. Testen er et redskap som kan gi kunnskap om hva barnet mestrer og ikke mestrer innenfor det aktuelle området, og i noen tilfeller om andre forhold av betydning for barnets læring og utvikling. Det kan for eksempel dreie seg om strategier og arbeidsstil, oppmerksomhet og utholdenhet. Pedagogiske tiltak og eventuell særskilt tilrettelegging kan utformes på basis av den innsikten som testresultatene gir. Ideelt sett bør testene, som et element i en mer helhetlig kartlegging, kunne bidra til å identifisere barnets nærmeste utviklingssone (Hagtvet, 2004). I klinisk sammenheng vil tolkning av testresultater ofte danne en viktig del av grunnlaget for beslutninger om videre utredning, og om art og omfang av spesialpedagogiske tiltak. Barnets behov for ekstra ressurser og særskilt tilrettelegging blir vurdert i lys av blant annet testresultater. Den betydning testresultatene tillegges kan med andre ord ha store konsekvenser for personen det gjelder. Testing har dermed en opplagt etisk dimensjon (Messick, 1980; Messick, 1989). Når tester brukes i forskning, er riktig tolkning av testresultater en forutsetning for forskningens vitenskapelige verdi. På grunnlag av tolkninger av testdata utvikles teorier og kunnskapsparadigmer, som i sin tur får konsekvenser for videre forskning så vel som pedagogisk og spesialpedagogisk praksis. Det er, for å sitere Befring, (2009, ikke sidetall) ”*selvinnlysende at forskning vil gjøre større skade enn gagn dersom den bygger på misvisende teoretiske premisser, mangelfulle og invalide data, feilaktige analyser eller lite troverdige vurderinger og konklusjoner.*”

Det er med andre ord opplagt viktig at den som bruker en test har grunn til å være rimelig sikker i sine tolkninger av testdataene. Hva slags informasjon ligger egentlig i testresultatet? Hva sier testskåren noe om? Å tolke et testresultat innebærer å tillegge resultatet en bestemt betydning eller mening. Det trekkes dermed visse slutninger på grunnlag av testresultatet. Det er disse slutningene som kan være mer eller mindre

gyldige, eller valide. Validitet refererer med andre ord til tolkningen av testresultatet, og ikke til selve testen eller skåren som sådan (Messick, 1989; Kleven, 2002a). Å tolke en testskåre vil ifølge Kane (1992) si å forklare hva skåren betyr. En slik forklaring må ifølge Kane (1992, s. 527) være "*an interpretative argument*", det vil si et resonnement som begrunner nettopp denne tolkningen av testresultatet. Kvaliteten på dette resonnementet blir dermed avgjørende for tolkningens troverdighet. Tolkningen kan aldri verifiseres i streng forstand, men det bør kunne dokumenteres at tolkningen har god støtte i tilgjengelig evidens. Et sentralt spørsmål er da hva som kan anføres som "evidens" i en slik sammenheng.

Validitet må alltid ses i relasjon til et konkret formål med å bruke en test. Det er som nevnt ikke testen som sådan, men slutninger som trekkes på grunnlag av testresultatet, som kan være mer eller mindre valide. Validitetsspørsmålet melder seg derfor strengt tatt først når det spesifiseres hva testresultatene skal brukes til, hva de er ment å si noe om (Fiske, 2002; Sireci, 2007). Det innebærer at en test kan være valid for ett formål, men ikke et annet. I "formål" ligger blant annet hvilken slutning man mener å kunne trekke på grunnlag av testresultatet, og i mange tilfeller også hvilken populasjon en slik slutning er gyldig for. Det vil si: Validitet må ses i sammenheng med *hva* man vil bruke testen til, og *hvem* man vil bruke den på. Viktige populasjonsparametre er blant annet alder, språklig-kulturell bakgrunn og klinisk gruppetilhørighet. Dette poenget kan utdypes med Bus Story som eksempel: I en undersøkelse av denne testens evne til å skille mellom barn med og uten spesifikke språkvansker, fant Pankratz og medarbeidere (2007) at barn med minoritetsbakgrunn var overrepresenterte i gruppen som feilaktig ble vurdert å ha språkvansker. Det forskerne kaller testens diagnostiske validitet viste seg således i denne undersøkelsen å være betinget av barnas språklig-kulturelle bakgrunn. Hvis vi holder oss til Bus Story, så må det prinsipielt også holdes åpent om testen kan brukes og tolkes på samme måte overfor norske barn som overfor engelske og amerikanske, og hvilken betydning barnas alder har. Man må for eksempel spørre hva fortrolighet med oppgaveformen kan bety for prestasjonene, og om det i så henseende kan være systematiske forskjeller mellom engelske og amerikanske barn på den ene siden og

norske barn på den andre. Det kan tenkes at ulikheter i art og omfang av språklig stimulering som barn i ulike land blir til del i hjemmemiljø, barnehage og førskole, har betydning for hvordan de mestrer en narrativ test. Det har igjen konsekvenser for hvordan resultatet kan tolkes. Tidligere skolestart blant engelske og amerikanske barn kan spille inn, derved at barn i disse landene får mer erfaring enn norske barn med den type oppgave som Bus Story representerer. På den annen side fant Aukrust og Snow (1998), som sammenlignet bordsamtaler i ulike land, at narrativ diskurs utgjorde en større del av bordsamtalene i norske hjem enn i amerikanske, hvor forklaringer sto mer sentralt. I alle tilfelle bør det reflekteres over hvilken betydning språklig og kulturell bakgrunn kan tenkes å ha for barns prestasjoner på en test.

Når det gjelder betydningen av klinisk gruppetilhørighet, for eksempel om skåren i prinsippet kan antas å utsi det samme for barn med og uten språkvansker, er det nærliggende å tenke at det ikke uten videre er tilfelle. For eksempel har flere forskere antydnet at barn med spesifikke språkvansker har svekkelser i den verbale korttidsminnefunksjonen (Leonard, 1998). Det er grunn til å tro at dette vil påvirke evnen til gjenfortelling. I så fall kan det tenkes at en lav skåre på en test som Bus Story for disse barna i større grad enn for barn med normal språkutvikling reflekterer minnefunksjoner, og i mindre grad språklige ferdigheter som ordforråd og grammatikk. Dette kan i hvert fall ikke utelukkes, selv om den støtten som bildene gir, antagelig bidrar til at minnefunksjoner får mindre betydning. I denne sammenheng kan også en undersøkelse av Miles og Chapman (2002) nevnes. De fant at barn med Down syndrom hadde bedre narrative ferdigheter enn barn med normal språkutvikling som var matchet på mer avgrensede tester for ekspressivt språk. Barna med Down syndrom uttrykte i en testsituasjon mer narrativt innhold enn kontrollgruppen. Forholdet mellom narrative ferdigheter på den ene siden og ekspressive semantiske og syntaktiske ferdigheter på den andre siden, var med andre ord annerledes blant barna med Down syndrom enn blant barna med normal språkutvikling. Miles og Chapman (2002) antyder at god syntaktisk forståelse og erfaring i å lytte til historier - altså fortrolighet med sjangeren - var viktige faktorer

som kunne forklare at barna med Down syndrom hadde bedre narrative ferdigheter enn det mer avgrensede mål for ekspressivt språk ga grunn til å forvente.

Denne diskusjonen illustrerer det generelle poenget om at testvaliditet må dokumenteres spesifikt med referanse til hvem som testes og hva som er testingens formål. Et viktig prinsipielt poeng er at det er problematisk å generalisere om testvaliditet på tvers av populasjoner. Det vil i praksis si, for eksempel, at selv om en test synes å ha visse egenskaper når den brukes på engelskspråklige seksåringer, kan vi ikke uten videre trekke den slutning at den fungerer tilsvarende for norske fireåringer.

Det nevnes ofte som en styrke ved narrative tester at de kan gi mye informasjon om barns språklige ferdigheter, og derfor har stor potensiell klinisk nytteverdi (Paul & Smith, 1993; Botting, 2002; Fey et al., 2004). Å fortelle eller gjenfortelle en historie krever koordinering av fonologiske, semantiske, morfosyntaktiske og pragmatiske ferdigheter. Prinsipielt kan man da tenke seg at et barns testrespons kan gi informasjon om alle disse aspektene ved språket. Oppgaven er språklig og kognitivt kompleks, og barnets aktivitet i testsituasjonen likner mer på normal språkbruk enn det som er vanlig ved en rekke andre språktester. Blant annet slike betraktninger ligger til grunn for anbefalinger om bruk av narrative tester for å kartlegge språkferdigheter (Bishop og Edmundson, 1987; Howlin & Kendall 1991; Paul & Smith 1993; Botting 2002; Fey et al., 2004). Det sies også at narrative ferdigheter har vist seg å kunne predikere språkutvikling, lese- og skriveferdigheter og skolefaglig suksess (Feagan og Applebaum, 1986; Paul & Smith 1993; Botting 2002). Pankratz og medarbeidere (2007) påpeker imidlertid at det er nødvendig å reflektere grundig over hva slags informasjon skårer på den enkelte narrative test faktisk er i stand til å gi. Uten slik refleksjon øker faren for at man legger for mye inn i et testresultat, og trekker slutninger som det ikke er grunnlag for. Pankratz og medarbeidere (2007) understreker at narrative tester kan brukes til ulike formål, og at validiteten må begrunnes for hvert enkelt formål. Deres artikkel kan ses som en advarsel mot å tillegge en test validitet for et bestemt formål uten at dette er spesifikt undersøkt, kun



basert på andre, ikke direkte relevante undersøkelser. Man bør etter disse forskernes oppfatning ikke legge til grunn en generell oppfatning om hva ”narrative tester” kan gi kunnskap om. Hver enkelt test må undersøkes for seg, og det må spørres helt spesifikt hva de enkelte skårer som beregnes kan antas å si noe om. Dersom prestasjonen kommer til uttrykk i form av en testskåre, må validitetsspørsmål rettes mot nettopp denne testskårens brukbarhet som grunnlag for bestemte slutninger. Hvilket grunnlag har man for å si at testskåren gir en bestemt type informasjon? Flere validitetsteoretikere framhever slike poenger. Kleven (2002a) peker på at validitetsspørsmålet alltid må ses i sammenheng med hva et måleresultat skal brukes til. Sireci (1998) understreker at dersom en test skal brukes til et bestemt formål, må det legges fram tilstrekkelige bevis for at nettopp denne bruken av testen er forsvarlig. Sireci (2007) minner også om at testvalidering ikke er noe som gjøres en gang for alle, men er en kontinuerlig prosess.

## 2.2 Innholdsvaliditet og begrepsvaliditet

Et sentralt spørsmål er altså hvordan man kan begrunne en påstand om hva et gitt testresultat innebærer. I en diskusjon av dette spørsmålet kan det være nyttig å ta utgangspunkt i begrepene innholdsvaliditet og begrepsvaliditet (content validity og construct validity), slik disse brukes innenfor testteori (Messick, 1980; Plante og Vance, 1994; Kleven, 2002a). **Innholdsvaliditet** foreligger ifølge Kleven (2002a) i den grad en måling (for eksempel et testresultat) faktisk gjenspeiler det atferdsområdet man ønsker å få vite noe om. ”Atferdsområde” brukes i generell betydning, og kan være for eksempel et sett av egenskaper eller et ferdighetsområde. I engelskspråklig litteratur tales ofte om det ”domenet” testen er ment å gi informasjon om. Termen ”konstrukt”, som også brukes, framhever at det man er interessert i, er et sammensatt fenomen og noe som ikke kan observeres direkte, og som i den forstand er en teoretisk konstruksjon. Innholdsvaliditet vurderes skjønnsmessig (Kleven 2002a). Utgangspunktet er da en teoretisk fundert oppfatning av hvilke elementer som inngår i atferds- eller kunnskapsområdet, og deretter en

skjønnsmessig vurdering av i hvilken grad testresultatene gir kunnskap om disse. Plante og Vance (1994, s. 15) taler om ”*expert judgment*” i denne sammenhengen. Testen bør dekke et rimelig stort og representativt utvalg av de elementene som inngår i det aktuelle ferdighetsområdet. Dersom en test for eksempel skal være et generelt mål for grammatisk forståelse hos barn over en viss alder, bør blant annet passivkonstruksjoner og ulike typer relativsetninger være elementer i testen. I motsatt fall kan man hevde at resultatet ikke er valid, fordi helt vesentlige aspekter ved ferdighetsområdet ikke dekkes. At testen fanger inn vesentlige sider ved ferdighetsområdet, kalles konvergerende validitet. Samtidig bør resultatet kun i liten grad påvirkes av andre forhold, som ikke anses som direkte relevante for det området man tester. Hvis for eksempel evne til logisk resonnering testes, bør testen være utformet slik at et godt resultat ikke forutsetter gode leseferdigheter. Eller for å holde oss til grammatikk eksemplet: En test av grammatisk forståelse hos barn bør konstrueres slik at størrelsen på ordforrådet ikke har betydning for resultatet (Bishop 2003/Lyster & Horn, 2009). Man må kunne utelukke at en lav skåre skyldes at barnet ikke forsto ordene som ble brukt i testen. At testresultatet ikke reflekterer faktorer utenfor det innholdsområdet man er interessert i, kalles diskriminerende validitet (Kleven, 2002a).

Bus Story ”tapper” høyst sannsynlig flere konstrukter. Narrativ produksjon og gjenfortelling er, som det ofte påpekes, en kompleks oppgave. Man kan gå ut fra at både språklige og kognitive ferdigheter testes. Det kan derfor ikke stilles som et ”krav” at den ikke tapper kognitive ferdigheter, eller sies at diskriminerende validitet er lav fordi kognitive evner spiller inn. Det er imidlertid viktig å vite noe om i hvilken grad testen måler språklige ferdigheter som ordforråd og grammatikk, og i hvilken grad den tester kognitive ferdigheter, som for eksempel hukommelse og generelt evnenivå. Dette er avgjørende for hvordan testresultatene kan tolkes. Dette illustrerer for øvrig poenget om at validitet alltid må relateres til et konkret formål med å bruke en test. Spørsmålet er: Hva ønsker vi informasjon om, og hvordan vet vi at testen er i stand til å gi slik informasjon?

Tankegangen som ligger til grunn for vurderingen av **begrepsvaliditet** er i vesentlig forstand den samme som ved innholdsvaliditet, men her kommer både teori og empiri inn i bildet (Kleven, 2002a). Begrepsvaliditet kan forstås som et uttrykk for i hvilken grad de relasjonene mellom målinger (i dette tilfelle testresultater) og ferdighetsområdet som man på teoretisk grunnlag antar er til stede, faktisk gjør seg gjeldende empirisk. Man kan undersøke dette ved å studere forholdet mellom testresultatet og andre målinger, nærmere bestemt målinger som testresultatet på teoretisk grunnlag forventes å stå i en bestemt relasjon til. Ut fra denne forståelsen kan begrepsvaliditet sies å være empirisk belagt innholdsvaliditet.

Plante og Vance (1994) representerer en tradisjon hvor det legges sterk vekt på at testvaliditet bør underbygges med empiriske undersøkelser. De gjør gjeldende at rent teoretiske vurderinger vil være subjektive og feilbarlige, og sjelden kan gi tilstrekkelig grad av sikkerhet for at testresultater faktisk reflekterer det innholdsområdet som testen er ment å dekke. Plante og Vance (1994) viser til at enkelte teoretikere av denne grunn bestrider at innholdsvaliditet kan anses som en form for validitet overhodet. Andre teoretikere har et mer ”positivt” syn på innholdsvaliditet. Sireci (1998) påpeker at rent teoretisk argumentasjon for at en test måler bestemte ferdigheter og egenskaper i mange tilfeller har tilstrekkelig overbevisningskraft. Det vil blant annet avhenge av hvor komplekst det aktuelle ferdighetsområdet er. For å relatere dette til språktester: Det er lettere å argumentere overbevisende rent teoretisk for høy testvaliditet hvis et resultat skal gjelde som et uttrykk for ”reseptivt vokabular” enn hvis resultatet skal gjelde som et uttrykk for ”språklige ferdigheter”, fordi konstruktet ”reseptivt vokabular” er lettere å definere og avgrense enn konstruktet ”språklige ferdigheter”. Sireci (1998) understreker videre at det er umulig både å konstruere en test og tolke testresultater uten at man på teoretisk grunnlag etablerer en sammenheng mellom testelementer og det underliggende innholdsområdet. Han påpeker at statistiske sammenhenger alene ikke kan påvise at elementene som inngår i en test er representative for det området som testes. Dersom innholdet i en test ikke kan *vurderes* som relevant for det aktuelle området (min utheving av Sireci 1998, s. 107: *”cannot be judged relevant to the*

*construct measured*”), så kan heller ikke statistiske sammenhenger gis en meningsfylt fortolkning. Av samme grunn er det også umulig å vurdere begrepsvaliditet uten å ta utgangspunkt i innholdsvaliditet. Det er jo nettopp en *antatt* sammenheng mellom testresultater og innholdsområdet, etablert på teoretisk grunnlag, som man ønsker å belegge empirisk. Plante og Vance (1994) foreslår selv at innholdsvaliditet kan konseptualiseres som et første skritt, som må følges opp av en empirisk bekreftelse. Kleven (2002a) er inne på det samme, idet han understreker at verken en rent rasjonalistisk eller en rent empiristisk tankegang er tilstrekkelig. Tilsvarende anser Sireci (2007, s. 481) at det å validere en bestemt bruk av en test ”*should involve both subjective analysis of test content and empirical analysis of test score and item response data.*” For å relatere dette til Bus Story: Hvis det finnes et godt teoretisk grunnlag for å anta at testen gir en god generell indikasjon på språklige ferdigheter, kan testen sies å ha høy innholdsvaliditet. Man kan så forsøke å underbygge validiteten empirisk gjennom å påvise signifikante og rimelig sterke statistiske sammenhenger mellom skårer på Bus Story og skårer på tester av ordforråd og grammatikk. Formålet med en slik empirisk bekreftelse er å frambringe bevis for at testen faktisk har de egenskaper som den på teoretisk grunnlag antas å ha. Dermed styrkes muligheten for valid bruk av testresultatene. Det etableres et sikrere grunnlag for å si hva som ligger i testskåren.

## 2.3 Strukturen i et validitetsargument – Toulmins modell

Den britiske vitenskapsteoretikeren Stephen Toulmin utga i 1958 boken ”The Uses of Argument” (Toulmin, 2003). Her tar han for seg hvordan man innenfor ulike fag og emneområder går fram for å begrunne påstander (”claims”), og hva som gjelder som akseptabel argumentasjon og bevisføring (Hitchcock & Veheil, 2006). Toulmin utviklet i denne boken en argumentasjonsanalytisk modell som siden har vært brukt i forskjellige faglige sammenhenger, med ulike modifikasjoner og nyanseringer (Hitchcock & Veheil, 2006). Toulmins modell har vært anbefalt som et egnet redskap i vurdering av testvaliditet generelt (Mislevy, 2002), og språktesters validitet spesielt

(Bachman, 2005). Bruken av Toulmins modell i denne sammenhengen bygger på en tilslutning til tanken om at enhver tolkning av et testresultat utgjør et resonnement, hvor en konklusjon - altså tolkningen - utledes fra et sett med premisser (Kane, 1992). Modellen kan brukes for å synliggjøre strukturen og innholdet i dette resonnementet. I modellen inngår følgende elementer:

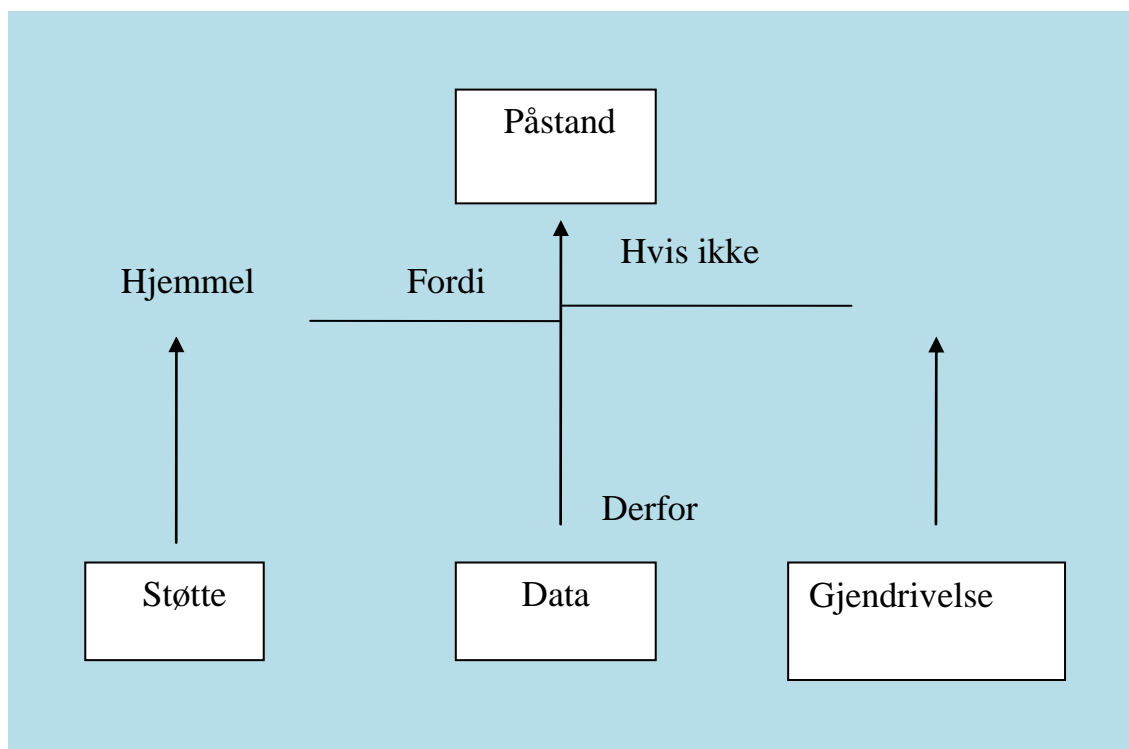
**Påstand** ("claim"): Dette er konklusjonen, eller tolkningen. Det vil i testsammenheng vanligvis si en påstand om at testpersonen besitter bestemte ferdigheter eller egenskaper i et visst omfang. Påstanden kan også sies å være svaret på spørsmålet: Hva forteller denne skåren oss? Eksempel: X har høy intelligens.

**Data**: Data er informasjonen som påstanden er basert på. I testsammenheng kan det for eksempel dreie seg om den skåren som testpersonen oppnår. Eksempel: X oppnådde skåren 116 på WISC-testen.

**Hjemmel** ("warrant"): Dette er generelle utsagn ("propositions") som brukes til å legitimere slutningen fra data til påstand. Eksempel: WISC-testen måler intelligens.

**Støtte** ("backing"): Støtte kan ses som svaret på spørsmålet om hvordan vi kan vite at den slutningen vi trekker når vi går fra data til påstand, er gyldig. Det vil i testsammenheng si: Hvordan legitimerer vi den slutningen vi trekker når vi går fra skåren til en påstand om hva skåren betyr? Uten tilstrekkelig støtte har hjemmelen ingen overbevisende kraft, og kan strengt tatt ikke kalles en hjemmel. Støtte kan utgjøres av teori, relevant tidligere forskning og/eller data som er særskilt frambrakt med det formål å validere testen. Eksempel: Her ligger i prinsippet all teoretisk og empirisk rationale bak en adekvat, autorisert tolkning av WISC-testen.

**Gjendrivelse** ("rebuttal"): Gjendrivelsen peker på grunner til at støtten ikke kan sies å være overbevisende. Dermed undergraves hjemmelen, og slutningen fra data til påstand framstår som tvilsom. Det vil også innebære at data tolkes annerledes. Eksempel: Her kan anføres enten kritikk av det teoretiske og empiriske rationale bak WISC-testen, eller spesifikke, situasjonsbestemte forhold som gjør tolkningen av testresultatet tvilsom. For eksempel: Testen måler i det alt vesentlige testpersonens utbytte av skoletilbudet han eller hun har hatt. Eller: Testleder ga for mye hjelp i testsituasjonen. Figur 2-1 viser Toulmins modell i diagramform.



Figur 2-1: Toulmins modell (etter Mislevy, 2003).

Resonnementet går fra data til påstand, via hjemmelen, som i sin tur bygger på ulike former for støtte. Gjendrivelsen kan peke på situasjonsbestemte forhold som gjør at slutningen fra data til påstand i et bestemt tilfelle ikke er gyldig, men kan også være teoretisk og empirisk kritikk rettet mot hjemmelen og støtten som sådan.

### 2.3.1 Vurdering av Bus Story-testens validitet i lys av Toulmins modell

Dersom Toulmins modell skal anvendes i en vurdering av Bus Story, må man først presisere hva testskårene eventuelt skal være valide mål for. Validitet må som nevnt demonstreres med henblikk på en bestemt bruk av testresultatene. I denne sammenhengen rettes oppmerksomheten mot problemstillingens to spørsmål: Med referanse til norske fireåringer spørres det om deres skårer på Bus Story er gode mål for språklige ferdigheter, og om skårene kan predikere videre språkutvikling. Det dreier seg da ikke om å vurdere påstander som noen faktisk har framsatt i akkurat denne formen. Påstandene er uttrykk for to mulige måter å bruke Bus Story-skårene på. Det er i seg selv interessant å undersøke om det faktisk er grunnlag for slik bruk.

Det kan samtidig sies at de to påstandene kan støtte seg på, eller er implisert i, utbredte oppfatninger om hva slags informasjon Bus Story-skårene er i stand til å gi. Som tidligere nevnt er det en vanlig oppfatning at narrative tester, herunder Bus Story, kan gi mye informasjon om barns språkferdigheter, og at Bus Story kan predikere språklig utvikling (Paul & Smith, 1994; Renfrew, 1997; Botting, 2002). Toulmins modell brukes her i en vurdering av om dette kan antas å gjelde også for norske fireåringer med normal språkutvikling. Under punktet ”støtte” anføres argumenter som kan tenkes brukt for å begrunne en slik antagelse. Under punktet ”gjendrivelse” blir disse argumentene problematisert. Aller først gis det imidlertid en litt fyldigere presentasjon av testen, for at diskusjonen skal bli lettere å følge.

Bus Story er en billedboktest som ble utviklet av Catherine Renfrew i 1969 (Renfrew, 1997). Ifølge manualen er den i første rekke en test av ekspressivt språk, men kan også gi informasjon om språklig forståelse. Testen er ikke offisielt oversatt til norsk. Bruken av Bus Story i Språkprosjektet bygger på den engelske manualen utarbeidet av Renfrew i 1997. Testen begynner med at testleder og barnet ser på billedboken sammen. Det er til sammen tolv bilder. Når barnet har gjort seg kjent med bildene, sier testleder: *”Nå skal jeg fortelle en historie om denne bussen (testleder peker). Når jeg er ferdig med å fortelle, vil jeg at du skal fortelle denne historien om bussen”*. Testleder skal holde seg strengt til historien slik den står i testmanualen, slik at alle barn får nøyaktig den samme historien. Tempo og pauser skal tilpasses i forhold til barnets alder og konsentrasjonsevne. Barnet promptes til å begynne sin gjenfortelling ved at testleder sier: *”Nå kan du fortelle historien om bussen til meg. Det var en gang...”*. Minimal og indirekte prompting er tillatt underveis for å drive gjenfortellingen videre, som for eksempel *”og så...?”*. Testen er ferdig administrert når barnet sier at det er ferdig med å fortelle.<sup>2</sup> Det gjøres lydopptak av gjenfortellingen. Før skåring skal både barnets gjenfortelling og testleders prompting transkriberes. Hvor barnets setninger starter og slutter skal markeres i transkripsjonen. Dette må tolkes ut fra holdepunkter som tonefall, pauser, innhold og

<sup>2</sup> Teksten til Bus Story er gjengitt i vedlegg 1. I vedlegg 2 gjengis en transkripsjon av en respons som ga en gjennomsnittlig skåre.

syntaks. Transkripsjonen utgjør skåringsgrunnlaget. Gjenfortellingen gis to ulike skårer, ved hjelp av et eget skåringsskjema som er en del av testmaterialet. Først skåres for informasjonsinnhold. Barnets setninger skåres med enten null, ett eller to poeng, ut fra i hvilken grad sentralt innhold er formidlet. Kun de av barnas setninger som er relevante for historien føres inn i skåringsskjemaet. Det er utarbeidet en guide som angir hvilke momenter som må med for å oppnå to poeng. Det kreves også at innholdsmomentene nevnes i korrekt rekkefølge. Setningslengde beregnes som gjennomsnittet av antall ord i de fem lengste setningene som det er gitt minst ett informasjonspoeng for, og det rundes av til hele tall. Innledende småord som *atte*, *også*, og *da* skal ikke medregnes, heller ikke gjentakelser av samme ord, uttrykk eller setning. Det kan her bemerkes at det i rammen av denne oppgaven ikke har vært mulig å undersøke hvor hyppig det har forekommet at barn har nevnt et innholdsmoment, men på feil sted, slik at det ikke har blitt poenggivende. Samtaler med testerne tyder imidlertid på at dette har skjedd i liten grad. Det vil si at lave informasjonsskårer alt overveiende gjenspeiler at innholdsmomenter ikke har vært nevnt overhodet.

### 2.3.2 Testskårenes validitet som uttrykk for norske fireåringers språklige ferdigheter

**Påstand:** *Fire år gamle X har gode språkferdigheter.* Med dette menes i denne sammenhengen at X har et godt ordforråd og gode grammatiske ferdigheter. Det vil si at ordforråd og grammatiske ferdigheter i denne sammenhengen tjener som operasjonelle variabler for det sammensatte konstruktet ”språkferdigheter”.

**Data:** *X oppnådde høye skårer på Bus Story.* Dette vil være basert enten på en subjektiv vurdering av at barnet gjorde det bra ut fra testadministrators erfaring, eller en sammenligning av skåren med normdata.

**Hjemmel:** *Skårene som fireåringer oppnår på Bus Story reflekterer i betydelig grad deres ordforråd og grammatiske ferdigheter.*



**Støtte:** Bus Story går ut på å gjenfortelle en historie. Dette er en kompleks språklig oppgave. Å gjengi narrativt innhold på en adekvat måte stiller ikke minst krav til ordforråd og grammatiske ferdigheter. Fordi både forståelse og produksjon er involvert i oppgaven, gjenspeiles både reseptive og ekspressive ferdigheter. Barnet må først oppfatte innholdet i fortellingen for å kunne gjengi det. Det betyr at barnet må forstå ordene og de morfologiske og syntaktiske betydningsselementene som forekommer i fortellingen. Med et godt utviklet ekspressivt ordforråd og gode ekspressive grammatiske ferdigheter øker også barnets muligheter for å gjenfortelle innholdet mest mulig komplett. Flere alternative måter å formidle innholdet på står da til rådighet for barnet, og han eller hun er mindre avhengig av å huske nøyaktig hvilke ord og uttrykksmåter testlederen brukte. Det er derfor grunn til å tro at barnas mestring av Bus Story vil gjenspeile deres ordforråd og grammatiske ferdigheter. Manualen til Bus Story (Renfrew, 1997) viser selv til en studie av Templin (1957), hvor gjennomsnittlig antall ord i de fem lengste setningene som barn ytret i en halvstrukturert lekesituasjon sammen med testlederen, ble funnet å være en god indikator på språklig utviklingsnivå. Ettersom barnet har bilder å støtte seg til i gjenfortellingen, stilles små krav til hukommelse. Det kan dermed anses som lite sannsynlig at det i vesentlig grad er hukommelsen som testes. I tillegg til dette tyder tidligere forskning på at Bus Story har god konvergerende validitet. I en amerikansk undersøkelse (Pankratz et al., 2007) viste det seg at skårer på tester av ordforråd og grammatikk kunne forklare en betydelig del av variansen i førskolebarns skårer på Bus Story. En test av ekspressive språkferdigheter (SPELT 3) viste en korrelasjon ( $r$ ) med Bus Storys informasjonsskåre på .68,  $p < .05$ , og med setningslengdeskåren på .57,  $p < .05$ . Den samme undersøkelsen fant også at barn med spesifikke språkvansker hadde signifikant lavere skårer enn barn med normal språkutvikling. Dette tolker forskerne som belegg for at Bus Story virkelig måler språklige ferdigheter.

**Gjendrivelse:** Selv om det generelt kan antas at evne til narrativ gjenfortelling er betinget av ordforråd og grammatikk, kan man likevel tenke seg at det er mulig å få høye skårer på Bus Story uten å ha spesielt godt ordforråd og gode grammatiske ferdigheter. I prinsippet kan man få full uttelling på informasjonsskåren ved å beskrive det man ser på bildene i et svært enkelt språk. I så fall er det ikke grunn til å

anta at informasjonsskåren gjenspeiler størrelsen på testpersonens ordforråd eller grammatiske ferdigheter. Det er heller ikke sikkert at setningslengdeskåren sier så mye om grammatisk kompetanse. Det er riktignok slik at gjennomsnittlig setningslengde, målt i morfer eller i ord, ofte brukes i forskning som et mål på språklig nivå (Brown, 1973). Dette går tilbake til Browns kjente verk fra 1973, *First Language*, hvor gjennomsnittlig ytringslengde, utregnet etter bestemte kriterier, ble funnet å være en god indikator for barns språklige utviklingsnivå. Men som Brown (1973) selv påpeker: Dette gjelder kun på et tidlig trinn i språkutviklingen, når ytringene er forholdsvis korte - Brown antyder selv fire morfer eller færre. Ifølge Brown (ibid) skyldes dette at ny grammatisk kunnskap på dette tidlige stadium i språkutviklingen nesten alltid gir seg utslag i økt ytringslengde. Senere bestemmes ytringenes kompleksitet - både syntaktisk og semantisk - i større grad av hva som blir sagt enn av ytringenes lengde. Det er således ikke sikkert at gjennomsnittlig setningslengde i Bus Story gjenspeiler grammatiske ferdigheter på en god måte. Ifølge den engelske normeringen av Bus Story er gjennomsnittlig antall ord allerede for fireåringene kommet opp i over fire (Renfrew, 1997). Ut fra Browns tankegang skulle da setningslengde ikke lenger være et godt mål for grammatisk utviklingsnivå. Man kan også spørre om det ligger en feilkilde i å telle ord i stedet for morfer. Dette kan illustreres med et eksempel: Ytringen ”*Buss kjøre der*” har tre ord og tre morfer.<sup>3</sup> Ytringen ”*Bussen kjørte der*” har også tre ord, men fem morfer. Den opplagte forskjellen i grammatisk kompleksitet mellom de to ytringene fanges ikke opp så lenge det er antall ord som utgjør skåringsgunnlaget, slik som i Bus Story.

Spørsmål om diskriminerende validitet må stilles både når det gjelder informasjon og setningslengde. Kan det tenkes at andre faktorer enn ordforråd og grammatikk har vel så mye å si? Kan for eksempel trygghet i situasjonen og selvtillit ha stor betydning for barnas prestasjoner? Dette kan både knyttes til personlighetstrekk som sjenanse og utadvendthet, og - muligens i kombinasjon med dette - fortrolighet med oppgaveformen. Det kan være store variasjoner med hensyn til hvor vant barn er med å uttrykke seg i en narrativ modus. Dette kan ha konsekvenser både for hvor

---

<sup>3</sup> ”Morf” betegner ”en minimal morfologisk enhet, det vil si en enhet som har uttrykk og betydning og som ikke kan deles i mindre enheter med uttrykk og betydning”, (Kristoffersen m.fl. 2003, s. 251).

komfortable de føler seg i testsituasjonen, og for hvor godt de forstår en slik oppgave som Bus Story utgjør. Hvordan vurderer barnet hva som er nødvendig å ta med i gjenfortellingen? Hvorfor utelater et barn informasjon? Er det fordi barnet ikke har oppfattet og forstått et bestemt innholdselement, eller kunne det ha nevnt det, men velger av en eller annen grunn å la være? Likeledes kan man spørre: Hvorfor uttrykker et barn seg i knappe ytringer? Skyldes det begrenset språklig uttrykksevne? Kan sjenanse eller utrygghet ha noe å si? Det er også grunn til å spørre om generelt evnenivå kan ha betydelig innvirkning på prestasjonen. Generelt evnerike barn kan for eksempel være bedre i stand til å oppfatte testlederens instruksjon, og ha bedre forståelse for hva oppgaven innebærer. Evne til desentrering kan her spille inn, det vil si evne til å vurdere hva som sett fra tilhørerens perspektiv utgjør en adekvat framstilling av historien. Karmiloff-Smith (1985) antyder således at barns narrative evner, blant annet utviklingen av overordnet strukturbevissthet, i betydelig grad gjenspeiler deres kognitive utvikling. Likeledes kan evne til å tolke bilder være en ferdighet som påvirker resultatet. I så fall er det også non-verbale evner som tappes, og ikke bare språklige ferdigheter. Det kan heller ikke utelukkes at minnefunksjoner spiller inn. Bildene gir riktignok en støtte som kan lette gjenkallingen, men historiens innhold og struktur framgår ikke klart av bildene alene. Man kan med andre ord tenke seg muligheten for at et velutviklet ordforråd og gode grammatiske ferdigheter verken er nødvendige eller tilstrekkelige betingelser for høye skårer på Bus Story.

I validitetsargumentet er det også vist til en undersøkelse som kan tenkes å støtte en slutning fra skårer på Bus Story til ordforråd og grammatikk. Pankratz og medarbeidere (2007) fant signifikante og rimelig sterke sammenhenger mellom Bus Story-skårene og andre språktester. Spørsmålet er imidlertid om disse resultatene kan overføres til norske fireåringer. Det bør aller først nevnes at Pankratz' undersøkelse er den eneste i sitt slag. Ingen andre undersøkelser har sett spesifikt på hvordan Bus Story korrelerer med andre språktester. Videre er det grunn til å merke seg at både barn med normal språkutvikling og barn med spesifikke språkvansker deltok i den nevnte undersøkelsen. 44 gutter og 20 jenter deltok. Barna var engelskspråklige, og deres gjennomsnittlige alder var fire år og ni måneder, altså nærmere fem enn fire.

Det er veldokumentert at det skjer mye med barns språklige utvikling mellom fire og fem års alder (Karmiloff & Karmiloff-Smith, 2001). Det er særlig grunn til å nevne at flere forskere taler om et kvalitativt sprang nettopp med hensyn til narrative ferdigheter når barn er mellom fire og fem år (Slobin & Berman, 1994; Karmiloff & Karmiloff-Smith, 2001). Deres narrative produksjon preges da ikke bare av mer avansert syntaks og rikere vokabular enn hos yngre barn, også de sjangertypiske elementene begynner å komme på plass. Barna synes altså å ha utviklet bedre forståelse for hva som kjennetegner den narrative sjanger. Det er da rimelig å anta at de er bedre i stand til å forstå hva en narrativ oppgave går ut på. På denne bakgrunn kan man spørre om en forutsetning for at Bus Story kan være en god indikator på språklige ferdigheter, er at oppgaven mestres rimelig bra. Kanskje er dette tilfelle først når barna har nådd en viss alder? Det er således ikke sikkert at ordforråd og grammatikk forklarer like mye av variasjonen i prestasjoner på Bus Story hos norske fireåringer som hos litt eldre amerikanske barn. Muligheten for at testegenskaper er relative til testpersonenes alder er absolutt til stede. Det kan tenkes at skårene på Bus Story gir en god indikasjon på språklige ferdigheter når barna er kommet i en viss alder, men at andre forhold enn språkevner i for stor grad innvirker på resultatet når barna er yngre. Da kan slikt som fortrolighet med og forståelse for oppgaven, grad av sjenanse, oppmerksomhet og konsentrasjon være viktige forklaringer på resultatet, det vil si en rekke forhold som ikke er direkte relevante for oppgaven, og som det er vanskelig å ha oversikt over. Det blir da problematisk å tolke resultatet som et uttrykk for språklige ferdigheter. Det bør endelig nevnes at Pankratz og medarbeidere (2007) fant at non-verbal intelligens forklarte en god del av variasjonen i barnas prestasjoner. Skårer på en test av non-verbal intelligens forklarte 18.4% av variasjonen på informasjonsskåren, og 11.1% av setningslengdeskåren. Dette indikerer ifølge forskerne at testen også gjenspeiler ikke-språklige ferdigheter.

Situasjonen synes altså å være: På den ene side kan det virke rimelig å anta at Bus Story gir en god indikasjon på fireåringers språkferdigheter. Et velutviklet ordforråd og gode grammatiske ferdigheter synes å være viktige forutsetninger for å mestre en narrativ gjenfortellingsoppgave. På den annen side kan det også argumenteres for at

andre faktorer enn ordforråd og grammatikk kan ha stor innvirkning på de skårene barna oppnår på Bus Story. Man kan tenke seg flere grunner til at fireåringer med godt ordforråd og gode grammatiske ferdigheter ikke oppnår høye skårer på Bus Story, og man kan tenke seg grunner til at fireåringer med svakt ordforråd og svake grammatiske ferdigheter oppnår relativt høye skårer. Diskusjonen kunne sies å vise at en teoretisk begrunnelse for høy validitet ikke er nok i denne sammenhengen. Den tolkningen av Bus Story-skåren som påstanden er et uttrykk for - altså at skårene gir en indikasjon på barnets ordforråd og grammatiske ferdigheter - synes å kreve sterkere, empirisk belegg. Det er nærliggende å vise til validitetsteoretikere som framholder at selv gode teoretiske argumenter for høy validitet ikke kan kompensere for svakt empirisk belegg (Plante & Vance, 1994).

### 2.3.3 Testskårenes validitet som prediktorer av norske fireåringers språkutvikling

I det følgende brukes Toulmins modell til å vurdere holdbarheten i en påstand om at norske fireåringers resultater på Bus Story gir grunnlag for slutninger om deres fremtidige språkutvikling.

**Påstand:** *Fire år gamle X vil sannsynligvis ha en god språkutvikling det neste året.*

For at denne påstanden skal kunne vurderes, må det presiseres hva som ligger i ”god språkutvikling”. Med dette menes i denne sammenhengen at X vil vise gode språklige ferdigheter om ett år, slik dette kan måles gjennom språktester. Tankegangen kan sies å være: Dersom Bus Story-skårene virkelig er gode prediktorer av språklig utvikling, bør det være en rimelig sterk sammenheng mellom skårene som fireåringene oppnår på Bus Story, og hvordan de gjør det på tester av ordforråd og grammatikk ett år senere.

**Data:** *X oppnådde høye skårer på Bus Story.*

**Hjemmel:** *Bus Story predikerer fireåringers språkutvikling på en god måte.*

**Støtte:** Som en narrativ test tapper Bus Story sannsynligvis et bredt spekter av språklige ferdigheter. Dermed gir en god indikasjon på hvordan språkutviklingen har

vært så langt. Da er det rimelig å anta at den også er i stand til å predikere framtidig språklig utvikling. Den oppfatning at Bus Story har god prediktiv verdi, støtter seg imidlertid i første rekke på tidligere undersøkelser hvor testen har vært brukt. Botting (2002, s. 2) hevder under henvisning til Stothard og medarbeidere (1998) at resultater på Bus Story har ”*strong relationships with future language and literacy performance*”. Pankratz og medarbeidere (2007) fant at Bus Story var en god prediktor av språk- og leseferdigheter hos barn med spesifikke språkvansker.

**Gjendrivelse:** Hvor godt testen kan antas å predikere fremtidig språkutvikling, avhenger av hvor godt den måler aktuelle språkferdigheter. Det kan da vises til diskusjonen under forrige punkt: Vi kan strengt tatt ikke vite hvor godt den måler aktuelle språkferdigheter, og rent teoretiske overveielser er ikke tilstrekkelige. Når det gjelder forskningen som det ofte vises til, er det nødvendig å se litt nærmere på hva de aktuelle undersøkelsene uttaler seg om. Bishop og Edmundson (1987) hadde som utgangspunkt en gruppe engelske fireåringer med spesifikke språkvansker. De samme barna ble testet halvannet år senere, og kunne da deles i to grupper: En hvor spesifikke språkvansker ikke lenger kunne påvises, og en hvor de fortsatt gjorde seg gjeldende. Vanskene ble på det grunnlaget karakterisert som henholdsvis forbigående (”*transient*”) og vedvarende (”*persistent*”). Undersøkelsen viste en tendens til at barna med vedvarende språkvansker hadde en spesielt lav skåre på Bus Story, nærmere bestemt på informasjonsdelen. Denne skåren kunne altså på en god måte predikere om barnas språkvansker var av forbigående eller vedvarende karakter. Forskerne konkluderer forsiktig: For fire år gamle barn med spesifikke språkvansker er prognosene bedre dersom de klarer å gjenfortelle innholdet i en enkel historie med støtte i bilder, enn hvis de ikke klarer dette. Dersom barna altså ikke mestrer denne oppgaven, er det mer sannsynlig at deres språkvansker er av vedvarende karakter, og ikke bare uttrykk for en forsinkelse i språkutviklingen. I en oppfølging av denne undersøkelsen fant Stothard og medarbeidere (1998) ytterligere indikasjoner på at Bus Story kan ha prediktiv verdi. Begge gruppene av barn, altså både ”vedvarende-gruppen” og ”forbigående-gruppen”, ble testet igjen i 15-16 års alder. Det viste seg da at noen av barna som hadde blitt plassert i ”forbigående-gruppen” faktisk hadde vedvarende, underliggende vansker. Disse vanskene hadde ikke blitt oppdaget da

barna hadde blitt retestet ved åtte og et halvt års alder (Bishop & Adams 1990). På det tidspunktet viste barna i ”forbigående-gruppen” ikke tegn til problemer med lesing og skriving, og de skåret normalt på fonologisk bevissthet. Men i 15-16-års alderen viste det seg altså at en del av disse barna likevel hadde vansker, som da kom til uttrykk i svake leseferdigheter og lave skårer på tester av fonologisk prosessering. Det er altså tale om barn som har overvunnet språkvansker og tilegnet seg et normalt språk mens de fortsatt er i førskolealderen, men som likevel har vedvarende, milde prosesseringsvansker som plasserer dem i en risikosone mht. utvikling av lese- og skriveferdigheter. Det interessante her er at denne gruppen viste seg å ha skåret relativt lavt på Bus Story da de var fem og et halvt. Det synes altså som om denne testen var spesielt sensitiv overfor deres vansker. Forskerne konkluderer med at barn som har diagnosen spesifikke språkvansker i fire års alder, men som synes å ha tilegnet seg normale språkferdigheter halvannet år senere, fortsatt kan være i risikozonen dersom de er relativt dårlige til å bruke sine formale språkferdigheter til å formidle informasjon i en narrativ oppgave, slik som Bus Story.

Disse undersøkelsene vies her en del plass, fordi henvisninger til nettopp disse ofte etterfølger nokså generelle utsagn om narrative testers eller ferdigheters prediktive egenskaper (Botting, 2002; Paul & Smith, 1993). Det presiseres ikke alltid at undersøkelsen konklusjoner gjelder bestemte forhold knyttet til språkutviklingen hos barn med spesifikke språkvansker. I en norsk masteroppgave hevdes det for eksempel at ”*Bishop og Edmundson (1987) fant at Bus Story med høy treffsikkerhet predikerte om 4-åringer ville ha språkvansker eller ikke ett og et halvt år senere*”(Nordeide, 2008). Men at Bus Story har slike prediktive egenskaper kan ikke sies å være påvist i Bishop og Edmundsons artikkel. Undersøkelsen fra 1987 tyder på at Bus Story kan være et nyttig redskap i utredning og kartlegging av spesifikke språkvansker. Den kan være egnet til å si noe om arten av spesifikke språkvansker hos fire år gamle engelskspråklige barn. Nærmere bestemt kan denne testen gi en indikasjon på om det virkelig dreier seg om spesifikke språkvansker av vedvarende karakter, eller om vanskene er av forbigående art, og altså mer uttrykk for en forsinkelse i språkutviklingen. For fireåringer med spesifikke språkvansker er det et

godt tegn dersom de har en relativt høy skåre på Bus Story. Det er da større sannsynlighet for at de senere vil skåre innenfor normalen på de øvrige språktestene. Bishop og Edmundsons (1987) funn kan støtte en antakelse om at vansker med å integrere språklig-kognitive delferdigheter for å formidle semantisk innhold berører kjernen i fenomenet spesifikke språkvansker. Dette er i tråd med flere forskeres teorier (Leonard, 1998). En gruppe teorier går ut på at spesifikke språkvansker skyldes redusert språklig-kognitiv prosesseringskapasitet (Leonard, 1998; Ellis-Weismer, 1999), og at vanskene manifesteres når denne kapasiteten overskrides. Det kan forklare en tendens til at barn med spesifikke språkvansker gjør det relativt dårligere på en kompleks test som Bus Story enn på deltester av ordforråd og grammatikk, noe som også er påvist i andre undersøkelser (Kendall & Howlin, 1991). Dette er interessante funn, men gir ikke grunnlag for generelle utsagn om Bus Story-testens evne til å predikere språkutvikling. Igjen er det nødvendig å spørre: Hvilken gruppe barn snakker vi om, hvor gamle er de, og hva er det testresultatet predikerer? Bishop og Edmundson (1987, s. 170) understreker for øvrig selv at resultatene fra deres undersøkelse har begrenset generaliseringsverdi ("*narrow applicability*"). De påpeker for eksempel at hos tre år gamle barn kan svak evne til gjenfortelling innebære noe helt annet enn for fireåringer ("*might have a very different significance*"). Vi kan også merke oss at det i disse undersøkelsene kun er informasjonsskåren som er funnet å ha prediktiv validitet. Det sies ikke noe om skåren for setningslengde.

Pankratz og medarbeidere (2007) konkluderer sin undersøkelse med at skårene på Bus Story synes å kunne predikere fremtidige språklige ferdigheter hos barn med spesifikke språkvansker rimelig godt. De sammenlignet et utvalg førskolebarns Bus Story-skårer med de samme barnas skårer på atten forskjellige språktester tre år senere. De fant statistisk signifikante sammenhenger mellom skårene på Bus Story og skårene på fjorten av de atten testene. For seks av testene var korrelasjonen (Pearsons  $r$ ) høy, for åtte moderat. I deres undersøkelse regnes  $r$ -verdier over .70 som høye, og  $r$ -verdier mellom .40 og .70 som moderate. Hvis vi vender tilbake til vårt utgangspunkt, og spør om disse funnene kan anføres som støtte for at Bus Story kan



---

predikere språkutvikling for norske fireåringer, må følgende bemerkes: Også denne prediksjonen gjaldt barn med spesifikke språkvansker. Barnas gjennomsnittsalder da de tok Bus Story-testen var 5;8, altså betydelig eldre enn fire år. Pankratz og medarbeidere (2007) påpeker selv at generalisering av disse funnene vanskelig gjøres av at utvalget var lite, og at det var skjevt mht. sosioøkonomisk status. Konklusjonen må bli at denne undersøkelsen bare i begrenset grad kan støtte antagelsen om at resultater på Bus Story kan predikere språklig utvikling for norske fireåringer.

## 2.4 Foreløpig oppsummering

Det er i det foregående foretatt en vurdering av Bus Story-testens validitet med utgangspunkt i Toulmins modell, det vil si testresultatenes validitet for nærmere bestemte formål. Diskusjonen gir grunnlag for følgende konklusjoner: På rent teoretisk grunnlag, inkludert henvisninger til empiriske undersøkelser med mulig overføringsverdi, kan det ikke argumenteres overbevisende for at de skårene som norske fireåringer oppnår på Bus Story gjenspeiler deres ordforråd og grammatikk, og i den forstand kan sies å være et valid uttrykk for deres språkferdigheter. Det kan heller ikke argumenteres overbevisende for at fireåringenes prestasjoner på Bus Story kan predikere deres språkutvikling særlig godt. Disse konklusjonene peker i retning av behovet for en empirisk undersøkelse av testens validitet med henblikk på nevnte forhold.

### 3. Metode

I følgende kapittel beskrives denne undersøkelsens metode, det vil si den framgangsmåte som er benyttet for å få svar på problemstillingen. Det innebærer en redegjørelse for oppgavens design, beskrivelse av utvalget av barn som er testet for å innhente data og prosedyrer ved datainnsamling, samt en beskrivelse av de testene som er brukt i denne oppgaven. Videre følger en vurdering av undersøkelsens validitet, herunder forhold som gjelder reliabilitet, og endelig belyses hvordan aktuelle forskningsetiske hensyn er ivaretatt i undersøkelsen.

#### 3.1 Design

I denne oppgaven foretas to ulike typer undersøkelser, i tråd med den distinksjonen mellom innholdsvaliditet og begrepsvaliditet som er redegjort for tidligere: På den ene siden en teoretisk fundert vurdering av hva slags informasjon skårene på Bus Story kan tenkes å gi (i den innledende teoridelen), og på den andre siden en empirisk undersøkelse av det samme. Det innebærer at en relativt fyldig teoridel, hvor også teoretiske drøftinger av problemstillingens to spørsmål inngår, danner bakgrunn for den empiriske undersøkelsen og drøftingen av data. Det er fremgangsmåten i den empiriske undersøkelsen som beskrives i dette kapitlet. Denne delen av oppgaven har et deskriptivt, ikke-eksperimentelt design (Kleven, 2002b), og er basert på analyser av kvantitative testdata. De to problemstillingene gjelder sammenhenger mellom undersøkte variabler. Sammenhengene blir utforsket og beskrevet ved hjelp av korrelasjonsstatistikk, og deretter drøftet nærmere. Testene som er brukt for å operasjonalisere undersøkelsens variabler, blir nærmere beskrevet i avsnittet om validitet. Først undersøkes de bivariate sammenhengene mellom Bus Story-skårene og skårene på testene for ordforråd og grammatikk. En sammenheng mellom to variabler  $x$  og  $y$  vil si en tendens til at en bestemt verdi på  $x$ -variabelen går sammen med en bestemt verdi på  $y$ -variabelen, og en annen verdi på  $x$ -variabelen går sammen med en annen verdi på  $y$ -variabelen (Heiman, 1996). I vår sammenheng vil det si

tendensen til at et bestemt resultat på Bus Story går sammen med bestemte verdier på testene for ordforråd og grammatikk, og at et annet resultat går sammen med andre verdier. En korrelasjonskoeffisient gir informasjon om styrken og retningen i denne tendensen. Fordi vi har data på forholdstallsnivå, brukes Pearsons  $r$  som korrelasjonsmål. I tråd med Heimans (1996) anbefaling presenteres korrelasjonsdataene også i form av spredningsdiagram. Spredningsdiagrammet gir et visuelt inntrykk av sammenhengenes styrke og form. Det er også interessant å undersøke hvor mye av variasjonen i barnas prestasjoner som kan forklares av forskjeller i ferdigheter innen ordforråd og grammatikk. Regresjonsanalyse benyttes for å kaste lys over dette spørsmålet. Det er en statistisk analyseteknikk som kan uttrykke hvor mye av variansen i en avhengig variabel som kan forklares av verdier på et utvalg uavhengige variabler (Lewis-Beck, 1980). Lewis-Beck (1980) påpeker at det må ligge teoretiske overveielser til grunn for hvilke uavhengige variabler man velger å inkludere. I denne oppgaven inkluderes de uavhengige variablene hvis forklaringskraft det er spesielt interessert i å undersøke, og som ut fra teoretiske overveielser antas å ha relativt stor betydning. Den primære interessen gjelder variablene ordforråd og grammatikk, fordi dette er å anse som språklige kjerneferdigheter (Bloom & Lahey, 1978). Det er som påpekt i teoridelen også grunn til å anta at kognitive ferdigheter kan ha betydning for hvor godt oppgaven mestres. Det er derfor av interesse å undersøke hvor store bidrag generelt evnenivå og hukommelse gir til å forklare variansen i barnas resultater på Bus Story.

Den andre problemstillingen gjelder hvorvidt skårene på Bus Story kan predikere språkutvikling, eller fremtidige språkferdigheter, hos norske fireåringer. Spørsmålet gjelder det som Kleven (2002a) kaller prediktiv kriterievaliditet, altså testens evne til å forutsi fremtidige prestasjoner. Spørsmålet belyses gjennom en undersøkelse av korrelasjoner mellom fireåringenes skårer på Bus Story, og deres skårer på et utvalg språktester administrert ett år senere. Disse skårene utgjør altså i denne sammenhengen operasjonaliseringen av begrepet ”språkferdigheter”. Hvis

signifikante og sterke sammenhenger kan påvises, tyder dette på at Bus Story har god prediktiv validitet.

## 3.2 Utvalg

Dataene stammer fra testing av en kohort uselekterte fireåringer fra en kommune i Akershus. Gruppen utgjør kontrollgruppen i en større studie av barn med forsinket eller atypisk språkutvikling. Barna er testet med et omfattende testbatteri, som gir informasjon om et bredt spekter av språklig-kognitive ferdigheter. Testingen ble i stor grad utført av mastergradsstudenter ved Institutt for Spesialpedagogikk, Universitetet i Oslo, med logopedi eller spesifikke læreversker som fordypningsfelt. Hver student testet cirka tjue barn. Prosedyrene ved testingen er nærmere beskrevet i avsnittet om reliabilitet. Utvalgskriteriene er ment å sikre at gruppen består av barn som fungerer innenfor normalområdet både språklig og kognitivt. Kommunen er valgt ut blant annet fordi den anses som representativ med hensyn til sosioøkonomiske variabler som inntekt og utdanningsnivå (<http://www.ssb.no/aarbok/emne05.html>).

Forskergruppen bak Språkprosjektet har kontaktet kommunen, som i neste omgang har sendt ut brev via barnehagene til aktuelle barn, eller direkte hjem til barn som ikke går i barnehage. Utvalgskriteriene er som følger:

Barn født i perioden 01.04.03 – 01.07.04, som ikke har kjente diagnoser som kan påvirke språkutviklingen, som ikke har utenlandsk navn, og/eller klart viser tilknytning til minoritetsspråklig gruppe, som ikke er henvist PP-tjenesten for språkvansker, som ikke får behandling for språkvansker av logoped eller audiopedagog, og som har norsk som morsmål.

Ved første kartleggingstidspunkt januar/februar 2008 var det registrert 201 barn i databasen. Barnas gjennomsnittsalder var da fire år og tre måneder. Opprinnelig var 250 barn forespurt. Svarprosenten er altså på ca. 80 %. Det er ingen indikasjoner på at frafallsgruppen avviker fra det endelige utvalget på relevante variabler. Frafallet

synes dermed ikke å representere en trussel mot utvalgets representativitet (Lund, 2002). Som nevnt er kommunen representativ for landet som helhet når det gjelder sentrale sosio-økonomiske bakgrunnsvariabler. Det er imidlertid grunn til å påpeke at kommunen i flere år har satset på et omfattende program for forebygging av lese- og skrivevansker. Tidlig kartlegging og oppfølging i barnehagen har vært et av tiltakene. Ifølge Stortingsmelding 16 (Kunnskapsdepartementet, 2006-07) har kommunens forebyggende innsats medvirket til at andelen elever med svake leseferdigheter på 2. trinn er betydelig redusert, og ligger på under det halve av landsgjennomsnittet. På dette grunnlaget kan det stilles spørsmål ved hvor representativt utvalget i Språkprosjektet er for norske fireåringer med hensyn til språklige ferdigheter. En mulig måte å undersøke dette på, er å se på hvordan barna i utvalget presterer i forhold til normdata der slike foreligger. Haakanes (2008), som undersøkte grammatiske ferdigheter hos disse barna da de var fire år, fant at de skåret høyere enn normgruppen på TROG-2, som er en test av grammatisk forståelse. Kan det tyde på at utvalget ikke er helt representativt? På den annen side skåret de lavere enn normgruppen på testen Grammatic Closure, som er en test av ekspressive grammatiske ferdigheter. En mulig tolkning av dette er at gruppen framstår som normal, for så vidt som de skårer til dels over, til dels under det som kunne forventes ut fra foreliggende normdata. Man kan imidlertid også spørre om den kommunale satsingen i større grad kan ha styrket utviklingen av reseptive ferdigheter, som TROG-2 tester, enn av ekspressive, som Grammatic Closure tester. I rammen av denne undersøkelsen er det ikke anledning til å gå grundigere inn på dette spørsmålet. Samlet sett synes det rimelig å legge til grunn at utvalget er representativt for norske fireåringer med normal språkutvikling med hensyn til språklige ferdigheter. For å styrke undersøkelsens validitet, er det hensiktsmessig å unngå for stor aldersmessig spredning. Jeg har derfor i denne oppgaven utelukket fra analysen barn som ikke har fylt fire år, og barn som er eldre enn fire og et halvt. Utvalget består etter dette av 184 barn, alle mellom 48 og 54 måneder ved første testtidspunkt. Gjennomsnittsalderen er 51 måneder. To barn har ”missing” på Bus Story. Det er uvisst om dette skyldes at barna ikke greide å gjennomføre testen, om det ble gjort feil i administreringen, eller andre årsaker. Analysene er kjørt på de data som ble gjort tilgjengelige for studentene

i februar 2009. Noen få barns resultater fra andre testtidspunkt var da av ulike grunner ennå ikke innført i datamatriksen. Utvalget ved andre testtidspunkt består derfor av 172 barn. Dette har ingen betydning for analysen, da utvalget fortsatt er stort, og det ikke er grunn til å anta at resultatene som ikke er kommet med, avviker vesentlig fra gjennomsnittet.

### 3.3 Tester brukt i denne undersøkelsen

I det følgende beskrives testene som er brukt i oppgaven. Testene blir her kun presentert. Hvorvidt de er egnet til å generere data som kan belyse oppgavens problemstillinger, det vil si om testene er gode operasjonaliseringer av de variablene som inngår i undersøkelsen, blir diskutert i avsnittet om validitet.

#### 3.3.1 Bus Story

Det henvises her til oppgavens teoridel (s. 23f), hvor Bus Story-testen er utførlig beskrevet.

#### 3.3.2 Bildebenevning (WPPSI-III+EVT 2)

For å måle ekspressivt vokabular ble deltesten Bildebenevning fra WPPSI-III (Wechsler, 2002) benyttet, supplert med åtte ekstra oppgaver fra Expressive Vocabulary Test, second edition (Williams, 2007). De ekstra oppgavene ble tatt med for å unngå takeffekt. Testen administreres ved at testleder viser barnet bilder av forskjellige objekter, og spør ”Hva er det?”, og barnet sier navnet på objektene. Testen består av 38 oppgaver med stigende vanskelighetsgrad.

### 3.3.3 Grammatic Closure

Grammatic closure er en deltest fra Illinois Test of Psycholinguistic Abilities (Kirk SA, McCarthy, Kirk WD, 1972). Den er en ekspressiv test med 33 oppgaver som først og fremst kartlegger morfologiske ferdigheter, men til en viss grad også syntaktiske. Barnet blir presentert for bilder (ett eller to) mens testleder sier en fullstendig setning, og en ufullstendig som barnet skal fullføre. For eksempel legges det fram to bilder, det ene viser en mann, det andre viser to menn. Testleder peker på bildene og sier: ”Her er en mann. Her er to...”, og så skal barnet si ”menn”. Oppgavene har stigende vanskelighetsgrad og testen avsluttes etter seks fortløpende feil. Grammatic closure ble normert til norsk i 1973-74.

### 3.3.4 British Picture Vocabulary Scale II (BPVS)

BPVS (Dunn LM, Dunn LM, Whetton & Burley, 1997) er utviklet for å måle barns ordforståelse. Den norske versjonen består av tolv sett med ord, og hvert sett inneholder tolv ord. Testen administreres ved at testleder viser barnet et ark med fire bilder, samtidig som et ord uttales. Barnet skal så velge ut det bildet som svarer til det aktuelle ordet. Ordene har stigende vanskelighetsgrad. Når barnet gjør åtte eller flere feil i et sett, stoppes testen. Testen er oversatt til norsk og standardisert for barn fra fire til seksten år av en gruppe forskere fra Institutt for Spesialpedagogikk, Universitetet i Oslo (Lyster, Rygvold & Horn, under utarbeiding).

### 3.3.5 Test for Reception of Grammar-2 (TROG-2)

TROG-2 (Bishop 2003/Lyster & Horn 2009) er utviklet for å teste grammatisk forståelse hos barn og ungdom. Den er konstruert slik at andre kilder til forståelsessvanser i liten grad skal kunne spille inn, som for eksempel svakt ordforråd og svakt korttidsminne. Testen består av åtti flervalgsoppgaver fordelt på tjue blokker, og hver blokk tester forståelsen av en spesifikk grammatisk konstruksjon. Barnet blir vist et ark med fire bilder for hver oppgave, og skal peke på det bildet som

svarer til en setning som testleder leser høyt. For å få poeng for blokken, må alle de fire oppgavene besvares korrekt. Det beregnes også en skåre for antall riktige enkeltoppgaver. Testen er oversatt til norsk og normert for aldersgruppen 4 til 16 år av en gruppe forskere ved Institutt for Spesialpedagogikk, Universitetet i Oslo.

### **3.3.6 Ordforråd (WPPSI-R)**

Resultater på de fire hittil nevnte testene brukes som indikatorer for ordforråd og grammatikk i forbindelse med oppgavens første problemstilling. Til den andre problemstillingen, som gjelder Bus Story-skårenes evne til å predikere språkutvikling, brukes de samme testene administrert ett år senere. I tillegg benyttes skårer på en test for dybdevokabular, Ordforråd fra WPPSI-R (Wechsler, 1989), som først ble administrert ved andre testtidspunkt, altså når barna er omkring fem år gamle. Testen begynner med tre billedoppgaver. Barnet skal navngi objekter som presenteres på bildene. Her gis ett poeng for hver riktig respons. Deretter følger verbaloppgaver, hvor barnet skal forklare begrepene som testleder presenterer muntlig. Barnets svar skåres med enten null, ett eller to poeng. Kvaliteten på definisjonen er avgjørende for poenggivningen. Generelt godtas enhver meningsfull forklaring på ordene og begrepene. For å få to poeng, kreves et godt synonym, angivelse av et vanlig bruksområde, en overordnet klassifikasjon, eller en korrekt beskrivelse av funksjon eller bruksmåte. Når det gjelder verb, må svaret inneholde et godt eksempel på handling eller årsakssammenhenger. Vage synonymer og mindre karakteristiske kjennetegn eller bruksområder er eksempler på svar som gir ett poeng. Barnet kan oppfordres til å utdype eller presisere svarene ved at testleder sier: ”Hva mener du?” eller ”Fortell meg mer om det?” Oppgavene har stigende vanskelighetsgrad, og testen avbrytes etter fem fortløpende feil.



### 3.3.7 Terningmønster

For å måle generelt evnenivå er testen Terningmønster fra WPPSI-III (Wechsler, 2002) benyttet. Testen er standardisert for barn fra 4 - 6;6 år. Barna skal sette sammen ulike mønstre med klosser, i begynnelsen etter modell av testlederen og etter hvert etter mønster fra bilder. Det er stigende vanskelighetsgrad på oppgaven.

### 3.3.8 Setningsminne

Testen Setningsminne er konstruert av forskerne bak Språkprosjektet med utgangspunkt i WPPSI-R (Wechsler, 1989) og Språk 6-16 (Ottem & Frost, 2005). Testen innledes med følgende instruksjon: ”Nå skal jeg si noe – og så vil jeg at du sier akkurat det samme. Si akkurat det samme som meg”. Barnet presenteres deretter for inntil 21 setninger med stigende vanskelighetsgrad. Helt korrekt gjengivelse gir ett poeng. Testingen avbrytes etter tre ufullstendige gjengivelser på rad. Testen gir en indikasjon på umiddelbar hukommelse for talt språk, som igjen antas å være influert av såvel kapasiteten i det språklige korttidsminnet som evnen til å hente semantisk, morfologisk og syntaktisk kunnskap fra langtidsminnet (Ottem & Frost, 2005).

## 3.4 Undersøkelsens validitet og reliabilitet

Validiteten i en forskningsundersøkelse refererer til gyldigheten i slutninger som inngår i undersøkelsen. For å vurdere validiteten i foreliggende undersøkelse, knyttes det an til Cook og Campbells validitetssystem (Shadish, Cook & Campbell, 2002).<sup>4</sup> Dette systemet er utviklet med tanke på undersøkelse av kausale forskningsproblemer i rammen av eksperimentelle design. Systemet omfatter fire kvalitetskrav som må være innfridd for at man skal kunne feste lit til slutningene i en kausal undersøkelse. De fire kvalitetskravene svarer til fire typer av validitet. Den foreliggende

---

<sup>4</sup> Kleven (2002a) understreker at Cook og Campbells system for vurdering av validitet i en undersøkelse ikke må forveksles med begrepsapparatet som brukes i forbindelse med testvaliditet (og som står sentralt i denne oppgaven). Kleven påpeker imidlertid at ”begrepsvaliditet” brukes likt i begge sammenhenger.

undersøkelsen har som nevnt et ikke-eksperimentelt design, og er basert på korrelasjoner. Det innebærer at Cook og Campbells system må tillempes noe. Ikke alle kvalitetskrav er like relevante, og kravene stilles på en litt annen måte. For eksempel er det vanlig i forbindelse med kausalundersøkelser å stille krav om at korrelasjoner må være signifikante og sterke, av hensyn til undersøkelsens såkalte statistiske validitet. I motsatt fall er den uavhengige variabelens effekt liten, og det blir problematisk å oppfatte den som en årsaksfaktor. I denne undersøkelsen er det ikke noe ”krav” om at sammenhengene må være sterke. Derimot er det viktig å kunne feste lit til at vi har et korrekt bilde av sammenhengene mellom variablene som inngår i undersøkelsene, uavhengig av om de er sterke eller svake. Dette forutsetter at visse forutsetninger er innfridd, noe som blir nærmere omtalt i avsnittet om statistisk validitet. Før validiteten i denne undersøkelsen diskuteres nærmere, blir forhold som angår datas reliabilitet behandlet. Denne rekkefølgen er valgt fordi reliabilitet er en forutsetning for validitet (Kleven, 2002a).

### **3.4.1 Reliabilitet**

Reliabilitet betegner i hvilken grad data er påvirket av tilfeldige målefeil (Kleven, 2002a), og er knyttet til nøyaktighet og presisjon i innsamlingen og bearbeidingen av data, i dette tilfelle testresultater. Hvis reliabiliteten er lav, kan man ikke gå ut fra at testresultatet faktisk gjenspeiler personens ferdigheter innenfor det området som testen måler, selv om testelementene i prinsippet er gode indikatorer. Dataene er da heller ikke egnet til å belyse undersøkelsens problemstilling. Reliabilitet er således en grunnleggende forutsetning for at data skal være brukbare, og dermed for hele undersøkelsens validitet. I det følgende diskuteres forhold med betydning for dataenes reliabilitet i denne undersøkelsen. Først omtales forhold knyttet til testadministrasjon og testbetingelser. Deretter omtales aspekter ved selve testene, herunder reliabilitetskoeffisienter.

En forutsetning for reliabilitet er at testene er administrert og skåret korrekt og enhetlig (Kleven, 2002a). I dette prosjektet har en rekke forskningsassistenter vært

involvert i testing og skåring, og ingen hadde betydelig testerfaring fra før.<sup>5</sup> Det oppstår dermed en fare for at inkonsekvent gjennomføring og skåring av testene setter preg på resultatene. For å unngå dette, fikk alle testerne grundig opplæring i administrering av testene og retningslinjer for skåring. Det ble også gjennomført praktiske øvelser hvor studentene fikk prøve materialet på hverandre. Underveis i testingen kunne prosjektlederne hele tiden kontaktes dersom noe var uklart. Ved noen anledninger ble testerne kalt inn til et felles møte for å drøfte problemstillinger som dukket opp i løpet av testperioden. Det ble da for eksempel avklart om et bestemt ord eller uttrykk skulle gjelde som poenggivende respons på en oppgave, eller hva slags prompting som kunne gis. Det ble også gjort lydopptak av de ekspressive testene, slik at tvilstilfeller kunne avklares i etterkant, og de prosjektansvarlige kunne kontrollere gjennomføringen av testene og forhold rundt testsituasjonen. God opplæring, mulighet for veiledning underveis, samt klare og standardiserte retningslinjer gir grunn til å anta at kravet om kompetent testadministrasjon og skåring er innfridd. Dette styrker tilliten til at dataene er reliable.

Rammene rundt testingen er også av betydning for datas reliabilitet (Kleven, 2002a). Det ble tilstrebet gunstige og mest mulig ensartede rammer omkring testingen. Det ble lagt vekt på skjermede og rolige omgivelser, og barna ble testet på omtrent samme tid på dagen, ikke under noen omstendighet sent. Testene ble alltid administrert i samme rekkefølge. Alle barn som gikk i barnehage ble testet der. Barn som ikke gikk i barnehage ble testet hjemme eller på det lokale PPT-kontoret. I utgangspunktet var barna alene med studenten under testingen, men om barnet ønsket det, kunne en forelder eller førskolelærer være til stede. Hvert barn ble testet over tre dager med en økt hver dag i første testrunde, og over to dager i andre testrunde. Hver økt varte i underkant av en time. Det bør bemerkes at enkelte testere i noen tilfeller opplevde støy fra naborom som sjenerende. Barna oppfattet da kanskje ikke instruksjoner eller promptinger slik som forutsatt, og det kunne være vanskelig for

---

<sup>5</sup> I denne oppgaven brukes data som er samlet inn på to forskjellige tidspunkt, når barna er fire og når de er fem. Nye studenter kom da inn i prosjektet. Beskrivelsen av prosedyrer for datainnsamling gjelder imidlertid begge tidspunktene. Forskjellen er kun at det var noen flere testere ved siste tidspunkt, og at testingen da gikk over kun to dager, da testbatteriet var noe mindre.

testleder å oppfatte barnets respons nøyaktig. Det kunne naturligvis være et problem ved tester hvor forskjellen mellom riktig og galt svar er en enkelt lyd eller et kort ord. Slike forhold ble imidlertid registrert og evaluert med tanke på resultatenes reliabilitet. Selv om ideelle rammer ikke alltid kunne oppnås, synes alt i alt hensynet til gode testbetingelser å være rimelig godt ivaretatt.

Når det gjelder de enkelte testene, kan det sies at BPVS, TROG-2 og Bildebenevning er enkle å administrere og skåre, noe også testernes tilbakemeldinger tydet på. Det samme gjelder Ordforråd (WPPSI-R), selv om det her må utøves noe mer skjønn i skåringen. Det viste seg at de standardiserte instruksjonene ved disse testene alt overveiende var tilstrekkelige til at barna forstod oppgavene og løste dem slik det var intendert. Dette er gunstig med hensyn til datas reliabilitet. For det første kan man da gå ut fra at barnet virkelig har forstått hva oppgavene går ut på. For det andre blir testledernes praksis mer enhetlig. Med Grammatic Closure forholdt det seg litt annerledes. Flere testere opplevde at en del barn ikke helt skjønnte oppgaven, og/eller ga svar som i og for seg ga god mening, men som ifølge manualen ikke skulle være poenggivende. Dette skapte usikkerhet når det gjaldt hvor mye prompting som skulle tillates, og når det gjaldt skåring. Testledernes praksis kunne her være litt forskjellig i starten. Det ble imidlertid raskt enighet omkring felles retningslinjer.

Reliabilitet ved språktestene samt testene Terningmønster og Setningsminne er også undersøkt ved hjelp av reliabilitetsmålet Cronbachs alfa. Alfakoeffisienten er et uttrykk for testens indre konsistens, det vil si hvorvidt de ulike delene av testen kan antas å måle det samme underliggende begrep. Alfaverdien er en funksjon av gjennomsnittlig korrelasjon mellom to og to oppgaver, og antall oppgaver i testen (Kleven 2002). Det siste gjenspeiler at tilfeldige feil jevner seg ut i det lange løp. Cronbachs alfa kan ha verdier mellom 0 og 1. Jo høyere verdi, desto bedre. Det er vanskelig å bestemme helt allment hva som er en akseptabel verdi. Innenfor samfunnsfagene er det vanlig å sette 0.70 som en omtrentlig grense, hvor alt over denne verdien normalt anses som tilfredsstillende (Ringdal 2007). Det er imidlertid nødvendig å ta hensyn til det konkrete forskningsspørsmål som står i fokus for å

avgjøre hva som er en akseptabel alfaverdi. Tabell 3-1 viser alfaverdiene for de ulike testene ved første testtidspunkt.

*Tabell 3-1: Alfaverdier T1*

<b>Tester</b>	<b>Alfaverdier T1</b>
Bildebenevning	.73
Grammatic Closure	.70
BPVS	.90
Trog 2-items	.97
Terningmønster	.54
Setningsminne	.68

Av tabell 1 framgår det at alfaverdiene er tilfredsstillende for alle språktestene, mens den er noe lav for Setningsminne, og i overkant lav for Terningmønster.

Terningmønster er likevel tatt med i den videre analysen. Den er brukt som en operasjonalisering av generelt evnenivå. Den lave reliabiliteten gir grunn til å spørre om resultater på denne testen virkelig er et valid mål i så henseende. Dette innebærer en usikkerhet som det må tas hensyn til ved tolkningen av resultatene.

Tabell 3-2 viser alfaverdiene for de ulike testene ved andre testtidspunkt.

*Tabell 3-2: Alfaverdier T2*

<b>Tester</b>	<b>Alfaverdier T2</b>
Bildebenevning	.74
Grammatic Closure	.70
BPVS	.92
Trog 2-items	.95
Ordforråd WPPSI	.71

---

Tabellen viser at alfaverdiene er tilfredsstillende for alle tester som er brukt ved andre testtidspunkt.

Det er en litt annerledes oppgave å vurdere reliabilitet i forbindelse med Bus Story-testen. De nevnte forhold knyttet til testadministrasjon og testbetingelser er naturligvis viktige på samme måte som ved de andre testene. Derimot lar det seg ikke gjøre å beregne en alfakoeffisient. Det kan også hevdes at skåringen av Bus Story i større grad enn ved de andre testene involverer skjønn og subjektive vurderinger (Nordeide, 2008). Den som skårer må blant annet vurdere om barnets ytringer dekker det aktuelle innholdet fullstendig eller bare delvis. Det er som nevnt også et krav at hendelsene skal presenteres i en bestemt rekkefølge. Dersom barnet presenterer en hendelse på feil sted i fortellingen, må det vurderes om plasseringen avviker så mye at det ikke skal gis poeng. Det vil ikke alltid være opplagte svar på slike spørsmål. Det knytter seg med andre ord en viss usikkerhet til vurderingen av barnas prestasjoner. Usikkerheten kan sies å være større enn ved testene for ordforråd og grammatikk. Slike forhold gir grunn til å spørre i hvilken grad de skårene barna oppnår er betinget av hvem som vurderer prestasjonene, med andre ord hvor god vurdererreliabiliteten er (Kleven, 2002). Dette var tema for en masteroppgave innenfor Språkprosjektet i fjor (Nordeide, 2008). Det ble her funnet at det var høy grad av samsvar mellom ulike skåreres vurderinger av de samme prestasjonene, og således konkludert med at vurdererreliabiliteten var god.

Alt i alt synes det rimelig å anta lav reliabilitet ikke representerer en alvorlig trussel mot denne undersøkelsens validitet. Den største usikkerheten knytter seg til testen Terningmønster. Alfaverdien kan her ikke sies å være tilfredsstillende, og konklusjoner som trekkes på grunnlag av data fra denne testen må tolkes med forsiktighet.

### 3.4.2 Validitet

#### *Innholds- og begrepsvaliditet*

I denne oppgaven reises spørsmålet om hvilken sammenheng det er mellom fire år gamle barns ordforråd og grammatikk og deres prestasjoner på Bus Story. Et første validitetsspørsmål som da kan stilles, er om det er benyttet gode indikatorer for ordforråd og grammatikk. Det vil i praksis si: Kan vi gå ut fra at de skårene barna oppnår på de aktuelle testene er valide uttrykk for deres ordforråd og grammatiske ferdigheter? Oppgavens rammer tillater ikke en grundig behandling av dette spørsmålet, men de viktigste momentene kan berøres. Utgangspunkt må tas i en definisjon av begrepene ordforråd og grammatiske ferdigheter. Det er lagt til grunn at reseptivt ordforråd betegner omfanget av ord som en person forstår, mens ekspressivt ordforråd betegner omfanget av ord personen er i stand til å bruke i sin egen tale. Reseptive grammatiske ferdigheter betegner evne til å forstå hvordan semantisk innhold uttrykkes gjennom morfologi og syntaks, mens ekspressive grammatiske ferdigheter betegner evne til å bruke morfologi og syntaks til å uttrykke semantisk innhold.<sup>6</sup> Spørsmålet er så om våre tester gir gyldig informasjon om disse evnene eller ferdighetene. Slik BPVS og Bildebenevning er utformet, er det etter min oppfatning plausibelt at barnas testresultater reflekterer størrelsen på henholdsvis reseptivt og ekspressivt ordforråd, under forutsetning av kompetent testadministrasjon og skåring. Dette er også tester som ofte brukes i forskning som indikatorer på ordforråd. TROG-2 tester forståelsen av et bredt spekter av grammatiske betydningselementer av både morfologisk og syntaktisk art, inkludert funksjonsord, pronomen, komparasjon, passiv og syntaktisk innføyning. Det er i utformingen av testen lagt vekt på at ordene som forekommer er slike som fireåring kan forventes å forstå (Bishop 2003/Lyster & Horn 2009). TROG-2 kan således anses å være en god indikator på grammatisk forståelse. Etter min oppfatning er imidlertid skårer på Grammatic Closure en mer diskutabel operasjonalisering av ekspressive grammatiske ferdigheter. Testen inneholder et nokså lite utvalg grammatiske strukturer, og krav til produksjon er begrenset. Det dreier seg stort sett om å fullføre

---

<sup>6</sup> Definisjonen er basert på, men ikke direkte hentet fra Enger og Kristoffersen (2002).

setninger ved hjelp av ett eller noen få ord. Videre er ord med uregelmessig bøyning overrepresentert, både når det gjelder substantiv, adjektiv og verb. Det kan da argumenteres for at det er barnets leksikon som tappes, ettersom slike bøyningsformer ikke kan utledes gjennom grammatiske regler, men må lagres som helheter i leksikon (Plunkett & Marchman, 1991; Simonsen, 2001). Man kan faktisk tenke seg at barn som gjør feil, såkalt ”gode grammatiske feil” (Hagtvet, 2004), er kommet lenger i utviklingen av grammatiske ferdigheter enn barn som svarer riktig: Et barn som sier ”bedre” eller ”skrev” henter - i hvert fall før morfologisk systeminnsikt er utviklet - en uanalysert form fra leksikon, mens et barn som sier ”godere” eller ”skrivet” bruker en grammatisk regel. I Språkprosjektet brukes ytterligere en test for ekspressive grammatiske ferdigheter, Past Tense-testen, som tester barns ferdigheter i fortidsbøyning. Man kunne tenke seg muligheten for å vurdere Grammatic Closure-testens validitet ved å undersøke hvor godt den samsvarer med Past Tense-testen, eller eventuelt bruke Past Tense-testen som indikator på ekspressive grammatiske ferdigheter. Problemet er da at vi måtte være sikre på at Past Tense-testen på sin side er bedre egnet til dette formålet. Imidlertid er det etter mitt syn like diskutabelt om Past Tense-skårene operasjonaliserer ekspressive grammatiske ferdigheter særlig godt. Bare elleve verb inngår i testen, og syv av disse er sterke. Dermed kan det også her antas at det i stor grad er leksikon som testes, og ikke morfologisk systeminnsikt. Jeg vil likevel legge til grunn at skårer på Grammatic Closure i denne sammenhengen er en akseptabel operasjonalisering av begrepet ekspressive grammatiske ferdigheter. Det bør bemerkes at de fleste sterke formene forekommer sent i testen, og at mange barn på grunn av stoppkriteriene ikke kommer til dem. Tidlig i testen dominerer svake, regelmessige former.

Oppsummerende kan det sies at tre av de fire språktestene er gode indikatorer for ordforråd og grammatikk, mens Grammatic Closure ikke er fullt så god. Det hefter derved en større usikkerhet omkring konklusjoner som baserer seg på denne testen. Endelig er det grunn til minne om, som Kleven (2002a) påpeker, at det i en forskningsundersøkelse aldri er tale om fullkommen validitet eller reliabilitet;



validitetssvikt og reliabilitetssvikt må påregnes i enhver undersøkelse, og er momenter det må tas hensyn til når resultatene tolkes.

Hensikten med å benytte testene Terningmønster og Setningsminne, er å få et inntrykk av i hvilken grad Bus Story stiller kognitive krav, utover de som allerede ligger i språktestene. Terningmønster, som er en av deltestene under utføringsdelen på WISC-III, er tatt inn i Språkprosjektet blant annet for å tjene som indikator for generelt evnenivå, og slik brukes den også i denne oppgaven. Det er en internasjonalt anerkjent test, og brukes ofte i forskning som operasjonell variabel, både for generelt evnenivå og non-verbal intelligens (f.eks. hos Stothard et al., 1998). Setningsminne gir en indikasjon på barnets umiddelbare hukommelse for talt språk (Ottem & Frost, 2005), noe som jo kan antas å være relevant i forbindelse med en gjenfortellingsoppgave slik som Bus Story. De to testene er egnet til å gi en indikasjon på i hvilket omfang Bus Story tapper kognitive ferdigheter som kan antas å ha betydning for prestasjonen på Bus Story. Som nevnt er imidlertid reliabiliteten for Terningmønster lav, og det gjør slutninger basert på resultater fra denne testen usikre.

### *Statistisk validitet*

Statistisk validitet i denne undersøkelsen refererer til gyldigheten i slutningene som trekkes angående de statistiske sammenhengene blant variablene i undersøkelsen. Den statistiske validitet kan sies å være høy dersom det er god grunn til å anta at vi har et korrekt bilde av sammenhengene, altså at styrken i disse verken over- eller undervurderes. Tre forhold er her av særlig betydning: Datas reliabilitet, utvalgets størrelse og representativitet, samt hvorvidt testresultatene er normalfordelte. Datas reliabilitet er vurdert, og er funnet å være god, med forbehold om at alfakoeffisienten for Terningmønster er noe lav. Undersøkelsen er basert på data fra et stort, representativt utvalg av individer fra den aktuelle populasjonen. Resultatene på de ulike testene er normalfordelte, noe som kommer til uttrykk i at verdiene for kurtosis

og skjevhet i hvert tilfelle ligger mellom -1 og 1. Det er således grunn til å anta at undersøkelsens statistiske validitet er tilfredsstillende.

### *Indre validitet*

Begrepet indre validitet i Cook og Campbells system refererer som nevnt til gyldigheten i slutninger som angår kausalforhold, og er knyttet til eksperimentelle design, hvor muligheten for å trekke slike slutninger er forholdsvis gode. Som ledd i undersøkelsen gis en påvirkning, og deretter undersøkes om påvirkningen har hatt noen effekt (Kleven, 2002b). I den foreliggende undersøkelsen studeres tingenes tilstand slik de er, og slik sett kan undersøkelsen kalles deskriptiv. Det som undersøkes, er sammenhengen mellom mål for ordforråd og grammatikk på den ene siden, og prestasjoner på den narrative testen Bus Story på den andre.

Sammenhengen styrke gir et svar på problemstillingen, som gjelder i hvilken grad skåren på Bus Story er en god indikasjon på språklige ferdigheter. Selv om undersøkelsen slik sett kan kalles deskriptiv, er antagelser om kausalforhold implisert, og det er derfor på sin plass å omtale noen forhold som kan sies å gjelde indre validitet. Den sammenhengen mellom for eksempel ordforråd og informasjonsskåre på Bus Story som avdekkes, antas å gjenspeile i hvilken grad ordforråd påvirker evnen til å formidle historiens innhold. Dette er en rimelig antagelse som kan støttes overbevisende rent teoretisk, og det grunnleggende retningsproblemet i forholdet mellom disse to variablene er lite problematisk.

Derimot kan man si at en variant av tredjevariabelproblemet gjør seg gjeldende: Kan det tenkes at andre variabler ligger ”bak” og påvirker prestasjonene både på Bus Story og på testene for ordforråd og grammatikk? Jeg vil her trekke fram tre forhold. De to første er generelt evnenivå og umiddelbar hukommelse for talt språk. Dette er variabler som vi kan kontrollere for statistisk, fordi det i testbatteriet inngår mål for disse egenskapene. En slik kontroll vil bli foretatt senere i oppgaven. Vi kan da få et svar på hvor mye generelt evnenivå og hukommelse for talt språk kan forklare av variasjonen i barnas prestasjoner, i tillegg til det som forklares av ordforråd og grammatikk. Det tredje forholdet, som vi ikke kan kontrollere for, kunne man kalle

fortrolighet med oppgaven, eller relevant språklig stimulering. Det kan tenkes at barnas prestasjoner på Bus Story i betydelig grad gjenspeiler hvor fortrolige de er med den narrative sjanger, hvor vant de er til å fortelle og bli fortalt for. Dette kan igjen knyttes til den språklige stimulering i hjemmemiljøet. Slik stimulering vil kunne påvirke ordforråd og grammatikk, og dermed fanges opp gjennom målene vi bruker for disse ferdighetene. Men man kan også tenke seg at ”fortrolighet med oppgaven” - uavhengig av ferdigheter innen ordforråd og grammatikk - i seg selv forklarer en del av variasjonen i barnas prestasjoner. Som nevnt er dette et spørsmål som det ikke er mulig å gi et svar på i rammen av denne undersøkelsen. Denne diskusjonen berører det grunnleggende spørsmålet om hva Bus Story egentlig måler, hvilke faktorer som forklarer barnas prestasjoner på denne testen. Et eksakt og utfyllende svar er det neppe mulig å gi. Det kan derfor være på sin plass å minne om at hovedformålet med oppgaven er å finne ut om og i hvilken grad skårene på Bus Story gjenspeiler ordforråd og grammatikk. Hvilke andre faktorer som også kan antas å ligge bak de skårene som barna oppnår, vil bli diskutert nærmere senere i oppgaven. Det blir da imidlertid tale om teoretiske antagelser og overveielser, og ikke empiriske undersøkelser (med unntak av generelt evnenivå og hukommelse for talt språk).

### *Ytre validitet*

Ytre validitet betegner i hvilken grad funnene fra undersøkelsen kan overføres til andre individer og settinger enn de som er undersøkt (Lund 2002). Denne undersøkelsen bygger på data fra et stort, representativt utvalg. Frafallet er på et akseptabelt nivå, og det er ikke grunn til å tro at frafallsgruppen avviker vesentlig fra det endelige utvalget på variabler som er relevante for undersøkelsen. Det innebærer at de sammenhengene som avdekkes i vårt materiale, neppe ville vært vesentlig annerledes med et mindre frafall. Det er derfor grunn til å tro at resultatene kan generaliseres til populasjonen som utvalget er trukket fra, altså norske fireåringer med normal språkutvikling. Det vil si at de konklusjonene som trekkes angående hva slags informasjon Bus Story-skårene kan antas å gi, har gyldighet når det er norske fireåringer med normal språkutvikling som testes. Det innebærer naturligvis også at

resultatene ikke kan generaliseres utover denne populasjonen. Hvilken informasjon som ligger i for eksempel seksåringers skårer på Bus Story, må eventuelt undersøkes for seg.

### 3.5 Etikk

Formelle etiske hensyn som frivillighet, anonymitet og informert samtykke er ivaretatt gjennom det overordnede prosjektet. Prosjektet er tilrådd av Personvernombudet for forskning ved Norsk Samfunnsvitenskapelig Datatjeneste og Regional komitè for medisinsk forskningsetikk. Ettersom prosjektet involverer testing av barn, er det tatt særskilte hensyn. Barn har særlige behov for og krav på beskyttelse (NESH, 2006; Befring, 2007). Barn bør bare involveres dersom undersøkelsen har klar nytteverdi, og det ikke medfører belastning å delta. Disse kravene vurderes som innfridd i Språkprosjektet. Språkprosjektets nytteverdi ligger i at det genereres økt kunnskap om norske barns språkutvikling, noe som er nødvendig for effektiv tidlig intervensjon og forebygging av språkvansker, så vel som generell stimulering og utforming av det pedagogiske tilbudet i førskole og skole. Barna har ikke selv gitt det formelle samtykket til å delta i testingen, selv om det naturligvis forutsettes at de er villige til å delta. Barna kunne, dersom de ønsket det, ha en voksen de kjenner til stede under testingen, for å dempe en eventuell belastning knyttet til testsituasjonen. En særskilt utfordring ved testing av barn gjelder nettopp spørsmålet om belastning ved å delta. Det er et selvsagt krav at testpersoner ikke skal oppleve ubehag og stress. Men barn kan ikke i samme grad som voksne forventes å si fra dersom de føler ubehag i testsituasjonen og ønsker å trekke seg. Ubegag kan for eksempel skyldes at barna blir slitne, eller at de føler at de ikke får til oppgavene. Testlederne er gjort kjent med sitt ansvar for å forsikre seg om at det enkelte barn ikke opplever testingen som en betydelig påkjenning, og for å avbryte dersom de har mistanke om at så er tilfelle. En slik inngripen må imidlertid bygge på testledernes subjektive vurderinger, og evne til å fange opp barnas signaler. Også dette kan være en evne som utvikles gjennom erfaring, og testerne i denne undersøkelsen har som

kjent begrenset erfaring fra før. Etiske problemstillinger, spesielt knyttet til barnas opplevelse av testsituasjonen og testledernes ansvar, ble derfor behørig tematisert i den opplæringen som testerne fikk. Mitt og andre testers klare inntrykk var at barna alt overveiende opplevde testsituasjonen som spennende og morsom, på tvers av variasjoner i utholdenhet, oppmerksomhet og konsentrasjon. Det synes alt i alt trygt å konkludere med at både formelle og uformelle etiske hensyn er godt ivaretatt i undersøkelsen.

## 4. Resultater

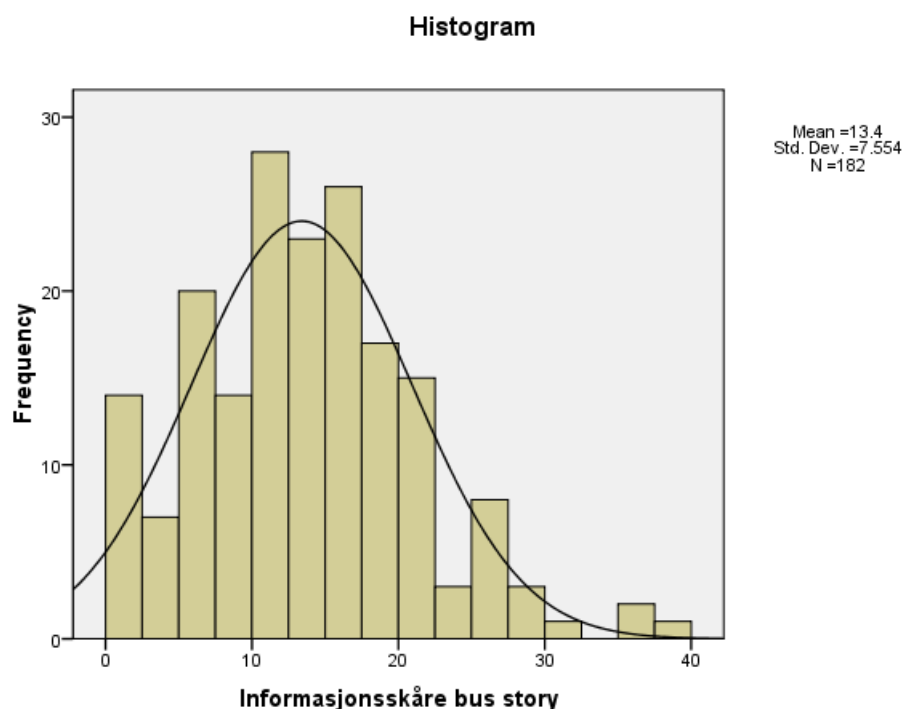
### 4.1 Forholdet mellom informasjonsskåren og setningslengdeskåren

Det viser seg at det er meget sterk korrelasjon mellom informasjonsskåren og setningslengdeskåren, med en Pearsons  $r$  på .81,  $p < .01$ . Dette tyder på at de to deltestene måler de samme språklige ferdighetene. Den høye korrelasjonen vitner om en sterk tendens til at de som klarer å formidle mye av historiens innhold, også bruker relativt lange setninger. Skåren for setningslengde kan ikke antas å gi ny informasjon i tillegg til det som ligger i informasjonsskåren. Det har i teoridelen vært pekt på visse problemer med å oppfatte setningslengde mer spesifikt som en indikator for grammatiske ferdigheter. Det er i utgangspunktet ikke helt klart hva testutvikleren anser at setningslengdeskåren gir informasjon om og hvordan den bør tolkes. Det er nærliggende å se skåren i forbindelse med bruken av gjennomsnittlig ytringslengde som et mål for grammatisk utvikling, slik dette ble introdusert av Brown (1973) - riktignok etter at Bus Story ble utviklet. Men som nevnt anser Brown (ibid) selv at sammenhengen mellom ytringslengde og grammatisk kompleksitet bare gjør seg gjeldende i barns tidlige språkutvikling, nærmere bestemt inntil gjennomsnittlig ytringslengde er på fire morfer eller mer. Det kan her også nevnes at det ved beregning av gjennomsnittlig ytringslengde anses nødvendig å legge til grunn langt flere ytringer enn det som gjøres ved skåringen av Bus Story. Ifølge Browns (ibid) egne regler skal det tas utgangspunkt i ett hundre ytringer. Endelig kan det nevnes at det i flere av de viktige undersøkelsene hvor Bus Story har vært brukt, kun er informasjonsskåren man har benyttet seg av (Bishop & Edmundsson, 1987; Stothard et al., 1998). På dette grunnlaget velger jeg i den empiriske undersøkelsen å konsentrere meg om informasjonsskåren. Dette gir også mulighet for en noe grundigere analyse. Når det i fortsettelsen refereres til resultater eller skårer på Bus Story, siktes det altså til skåren for informasjon.

## 4.2 Fireåringenes resultater på Bus Story

I følgende avsnitt gis en deskriptiv framstilling av barnas resultater på Bus Story.

Figur 1 viser variasjonen i fireåringenes resultater.



*Figur 4-1: Histogram som viser fordelingen på Bus Story*

Det er forholdsvis stor spredning i barnas prestasjoner. Gjennomsnittlig skåre er 13.5 poeng, mens standardavviket er på 7.5. Kvartildifferansen er på 9 poeng. Det innebærer at halvparten av barna skårer et sted mellom 9 og 18 poeng, mens en fjerdedel skårer under 9 poeng og en fjerdedel skårer over 18 poeng ( $13.5 \pm 4.5$ ). Det kan diskuteres om de høyeste skårene burde behandles som uteliggere og fjernes fra utvalget. Argumentet ville være at få enheter med svært høye verdier bidrar til å gi et skjevt bilde av spredningen i utvalget. Den høyeste verdien er 38, hvilket er cirka tre standardavvik over gjennomsnittet. Det er imidlertid flere barn som ligger opp mot denne ekstremverdien, så vedkommende ligger ikke helt "alene". Det gir etter min oppfatning det mest valide uttrykk for utvalgets prestasjoner å inkludere samtlige barn.

### 4.3 Sammenhenger mellom Bus Story og andre språktester

Tabell 4-1 viser hvordan Bus Story korrelerer med testene for ordforråd og grammatikk.

*Tabell 4-1: Korrelasjoner mellom Bus Story informasjon og andre språktester administrert ved fire års alder.*

Tester	Bildebenevning	Grammatic Closure	BPVS	Trog 2-items
<b>Bus Story</b>	.374**	.251**	.288**	.274**

\*\* . Korrelasjonen er signifikant på 0.01-nivå.

Det er signifikante, positive korrelasjoner mellom Bus Story og alle de fire testene for ordforråd og grammatikk. Det vil si at det er en sammenheng mellom hvordan barna gjør det på Bus Story-testen og deres prestasjoner på de andre språktestene.

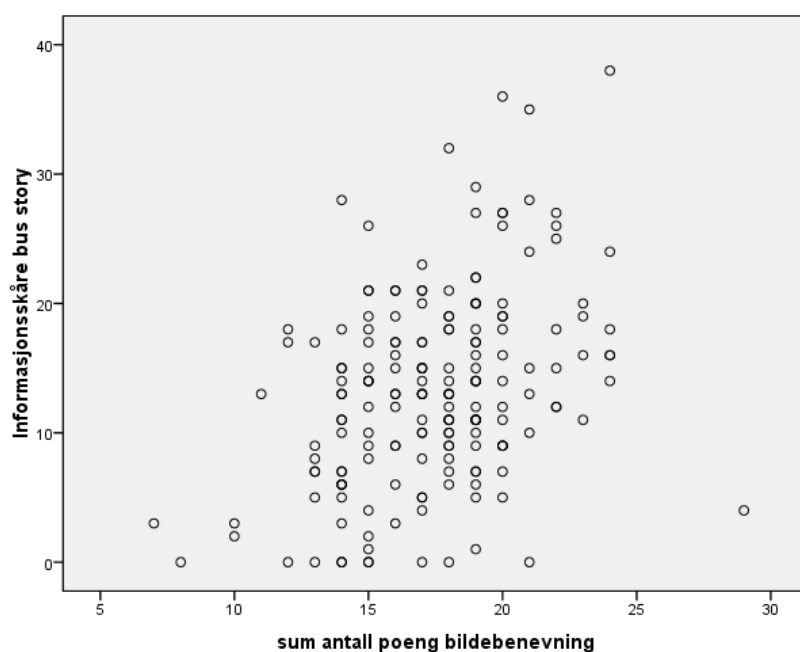
Korrelasjonskoeffisienten uttrykker hvor sterk sammenhengen er, nærmere bestemt hvor sterk tendensen er til at høy skåre på den ene testen går sammen med høy skåre på den andre, og lav skåre på den ene går sammen med lav skåre på den andre.

Sammenhengene kan beskrives som svake. Terminologien er riktignok ikke helt ensartet når det gjelder hva som betegnes som henholdvis svake, moderate og sterke sammenhenger. Denne oppgaven følger samme konvensjon som Pankratz og medarbeidere (2007) gjør i sin studie av Bus Story. De anser r-verdier mellom .70 og .90 som høye, mens verdier mellom .40 og .70 anses som moderate. Ut fra denne terminologien er alle sammenhengene som er avdekket i denne undersøkelsen svake. Alle r-verdiene er under .40. Bare mellom Bus Story og Bildebenevning nærmer r-verdien seg moderat, idet  $r=.374$ ,  $p<.01$ . Pankratz og medarbeidere (2007) betegner r-verdier over .70 som uttrykk for en *markert* sammenheng mellom variablene, og r-



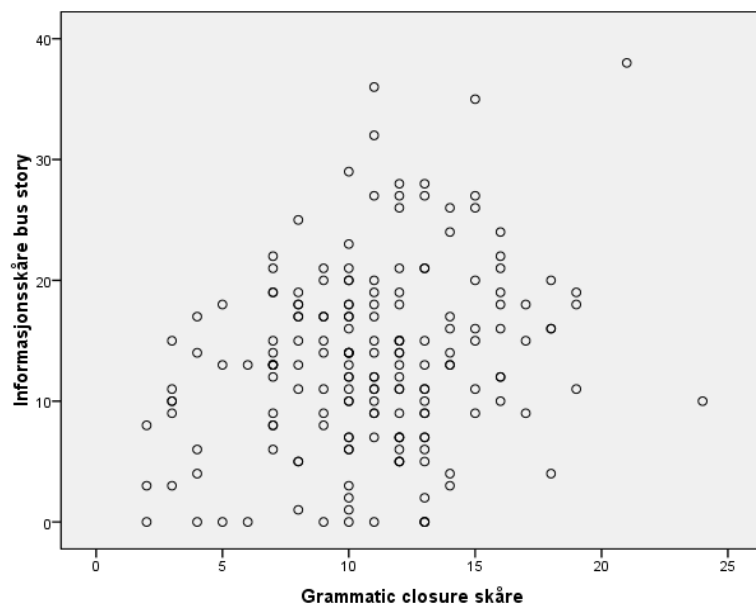
verdier over .40 som uttrykk for en *substansiell* sammenheng. Verdier under .40 gir ingen egen betegnelse. Det er kanskje ikke overraskende at den høyeste korrelasjonen finnes mellom Bus Story og Bildebenedvning, ettersom de begge er ekspressive tester. På den annen side må det da bemerkes at korrelasjonen mellom Bus Story og Grammatic Closure, som også er en ekspressiv test, er den svakeste av de fire.

For å gjøre det lettere å danne seg et inntrykk av tendensen til at bestemte verdier på Bus Story går sammen med bestemte verdier på de andre testene, presenteres spredningsdiagrammer for de bivariate sammenhengene. Heiman (1996) anbefaler at spredningsdiagram alltid presenteres sammen med korrelasjonskoeffisienter. Spredningsdiagrammet gir et visuelt bilde av sammenhengens styrke og retning. Ved sterke positive sammenhenger vil enhetene plassere seg omkring diagonalen som går fra lave verdier på begge aksene til høye verdier på begge aksene. Jo svakere sammenhengen er, desto mindre tydelig vil denne diagonalen avtegne seg. Figur 4-2 viser spredningsdiagram for sammenhengen mellom Bus Story og testen for ekspressivt ordforråd, Bildebenedvning.



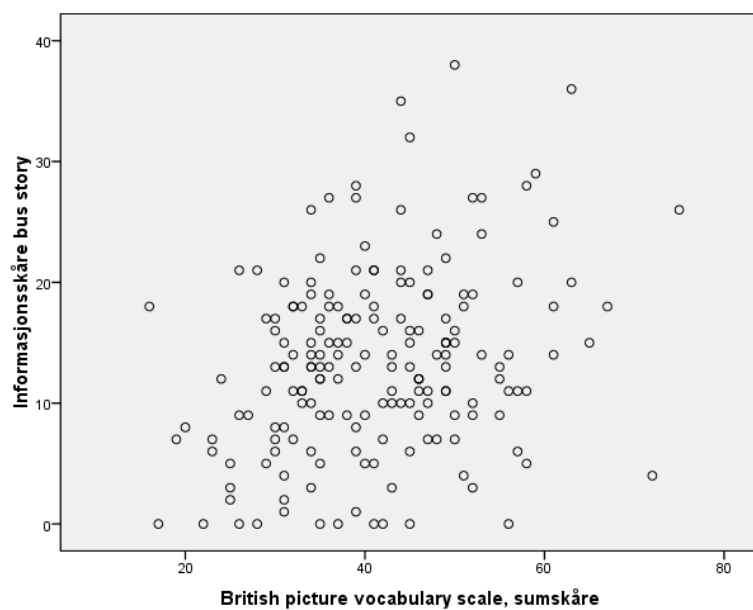
Figur 4-2 - Spredningsdiagram for sammenhengen mellom Bus Story og Bildebenedvning.

Figur 4-3 viser spredningsdiagram for sammenhengen mellom Bus Story og testen for ekspressive grammatiske ferdigheter, Grammatic Closure.



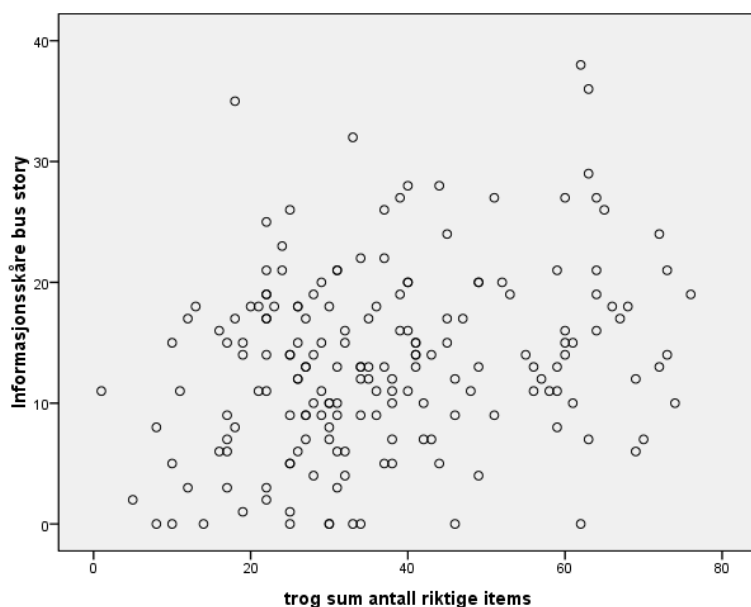
*Figur 4-3 – Spredningsdiagram for sammenhengen mellom Bus Story og Grammatic Closure.*

Figur 4-4 viser spredningsdiagram for sammenhengen mellom Bus Story og testen for reseptivt ordforråd, BPVS.



*Figur 4-4 – Spredningsdiagram for sammenhengen mellom Bus Story Informasjon og BPVS.*

Figur 4-5 viser spredningsdiagram for sammenhengen mellom Bus Story og testen for grammatisk forståelse, Trog-2.



*Figur 4-5: Spredningsdiagram for sammenhengen mellom Bus Story og Trog-2.*

Ut fra spredningsdiagrammene er det rimelig å opprettholde betegnelsen av sammenhengene som svake, med den nyansering at sammenhengen mellom Bus Story og Bildebenevning ligger nokså tett opp mot det som kalles moderat. Det er, for å bruke Pankratz og medarbeideres (2007) terminologi, ikke noen substansiell sammenheng mellom resultatene på Bus Story og noen av testene for ordforråd og grammatikk.

#### 4.4 Sammenhenger mellom Bus Story og kognitive tester

Som nevnt i teoridelen er det grunn til å tro at Bus Story måler ikke bare språk, men også visse kognitive ferdigheter. Det er derfor beregnet sammenhenger mellom resultater på Bus Story og på tester for generelt evnenivå og hukommelse for talt språk. Tabell 4-2 viser disse sammenhengene.

Tabell 4-2: Korrelasjoner mellom Bus Story og Terningmønster og Setningsminne

Tester	Setningsminne	Terningmønster
Bus Story	.297**	.178*

\*\* . Korrelasjonen er signifikant på 0.01-nivå.

\* . Korrelasjonen er signifikant på 0.05-nivå.

Tabell 4-2 viser at det er signifikante, positive sammenhenger mellom barnas resultater på Bus Story og på testene Setningsminne og Terningmønster.

Sammenhengene er sterkest mellom Bus Story og Setningsminne, med  $r=.297$ ,  $p < .01$  og svakere mellom Bus Story og Terningmønster, med  $r=.178$ ,  $p < .05$ . Begge sammenhengene er svake.

## 4.5 Språklige variablers bidrag til å forklare variasjonen i resultater på Bus Story

For å undersøke i hvilken grad språklige ferdigheter forklarer resultatet på Bus Story, er det gjennomført regresjonsanalyser med Bus Story som avhengig variabel, og de ulike språktestene som uavhengige variabler. Det kan her bemerkes at de innbyrdes korrelasjonene mellom de ulike språktestene er moderate, varierende fra  $r=.367$ ,  $p < .01$ , til  $r=.585$ ,  $p < .01$ . At sammenhengene er moderate, viser at testene måler ulike sider ved barnas språkferdigheter. Det gjør det interessant å undersøke hvor stor forklaringskraft hver enkelt variabel har. Ettersom Bus Story er en ekspressiv test, kan det forventes at tester av ekspressivt språk vil forklare den største andelen av variansen i barnas resultater. Det er derfor foretatt en regresjonsanalyse hvor de ekspressive testene er ført inn først (modell 1). Det er også et åpent spørsmål om Bus Story først og fremst måler barnas ordforråd, og om grammatiske ferdigheter har

noen selvstendig forklaringskraft utover det som testene for ordforråd kan forklare. Rekkefølgen er derfor endret i modell 2, hvor testene for ordforråd er introdusert først. Tabell 4-3 viser resultatet av disse regresjonsanalysene.

*Tabell 4-3: Hierarkiske regresjonsanalyser med Bus Story som avhengig variabel og Bildebenevning, Grammatic Closure, BPVS, Trog 2-items som uavhengige variabler.*

Steg	Variabel	R square	R square change	Sig.
Modell 1				
1.	Bildebenevning	.140	.140**	.000
2.	Grammatic Closure	.149	.009	.168
3.	BPVS	.155	.007	.242
4.	Trog 2-items	.166	.010	.144
Modell 2				
1.	Bildebenevning	.140	.140**	.000
2.	BPVS	.148	.009	.181
3.	Grammatic Closure	.155	.007	.224
4.	Trog 2-items	.166	.010	.144

\*\* . Korrelasjonen er signifikant på 0.01-nivå.

Av tabell 4-3 fremgår det at Bildebenevning forklarer 14% av variansen i resultater på Bus Story når den innføres først. De andre testene forklarer da ingen signifikant del utover dette. Dette gjelder hva enten den ekspressive testen Grammatic Closure eller ordforrådstesten BPVS kjøres inn i analysen i et andre steg etter Bildebenevning. Det innebærer at de tre testene ikke forklarer noe av variansen utover det som Bildebenevning har forklart. En forklaring på dette ligger sannsynligvis i at det disse testene måler, delvis er inkludert i Bildebenevning. De

språkferdighetene som de i tillegg måler, kan ikke bidra ytterligere til å forklare variansen i resultater på Bus Story. Samlet forklart varians av språktestene er altså på 14%. Dette må karakteriseres som overraskende lavt.

#### 4.6 Kognitive variablers bidrag til å forklare variasjonen i resultater på Bus Story

Det er som tidligere nevnt viktig i tolkningen av barns resultater på Bus Story å ha kunnskap om hvor stor innvirkning andre ferdigheter enn rent språklige kan tenkes å ha på resultatet. Generelt evnenivå og umiddelbar hukommelse for talt språk er blant de ferdighetene som kan forventes å ha en viss betydning. Det er derfor foretatt regresjonsanalyser med Bus Story som avhengig variabel og testene Setningsminne og Terningmønster som uavhengige variabler. Det er også undersøkt hvor mye disse to testene kan forklare etter at det er kontrollert for betydningen av språklige variabler, det vil si når de kjøres inn i analysen etter Bildebenevning (modell 3). Tabell 4-4 viser resultatet av disse analysene.

*Tabell 4-4: Hierarkiske regresjonsanalyser med Bus Story som avhengig variabel og Setningsminne, Terningmønster og Bildebenevning som uavhengige variabler.*

Steg	Variabel	R square	R square change	Sig.
<b>Modell 1</b>				
1.	Setningsminne	.088	.088**	.000
2.	Terningmønster	.110	.021*	.041
<b>Modell 2</b>				
1.	Terningmønster	.032	.032*	.016
2.	Setningsminne	.110	.078**	.000
<b>Modell 3</b>				
1.	Bildebenevning	.143	.143** <sup>7</sup>	.000
2.	Setningsminne	.180	.037**	.006
3.	Terningmønster	.189	.009	.157

\*\* Korrelasjonen er signifikant på 0.01-nivå.

Tabellen viser at Setningsminne og Terningmønster til sammen kan forklare 11% av variansen i resultater på Bus Story (modell 1 og 2). Av modell 1 og 2 framgår det at Setningsminne gir et større unikt bidrag enn Terningmønster. Når Setningsminne innføres i et steg nummer to etter Terningmønster, forklarer den 7.8% av variansen. Når Terningmønster innføres i et steg nummer to etter Setningsminne, forklarer den 2.1%. Av modell 3 framgår det at det kun er Setningsminne som kan gi en ytterligere forklaring på variansen i barnas resultater utover det som Bildebenevning allerede har forklart. Setningsminne forklarer 3.7% av variansen når den kjøres inn i analysen i et steg nummer to etter Bildebenevning. Terningmønster forklarer da ikke noe i tillegg.

<sup>7</sup> Den lille endringen i R square kan muligens forklares av en s.k. "suppression variable". Det kan være korrelasjonene mellom variablene som skaper en slik endring (Cohen, 2003).

Terningmønster kan heller ikke forklare noen signifikant del av variansen når den innføres som et steg to rett etter Bildebenevning (ikke vist i tabellen;  $r^2 \text{ change} = .008$ ,  $p > .05$ ). Resultatene som er framstilt i Tabell 5 tyder på at de kognitive kravene i Bus Story er små utover de kravene som allerede ligger i den ekspressive testen Bildebenevning. Terningmønster kan ikke forklare noen ekstra del av variansen i barnas testresultater, mens Setningsminne forklarer en liten, men signifikant del.

#### 4.7 Bus Story-testens prediktive validitet: Sammenhenger mellom Bus Story og språktester administrert ett år senere

Undersøkelsen så langt har vist svake sammenhenger mellom ordforråd og grammatikk og resultater på Bus Story. Det ser altså ut til at Bus Story-testen i beskjeden grad måler ordforråd og grammatiske ferdigheter. Som Kleven (2002a) påpeker, er det rimeligvis en sammenheng mellom hva en test måler og hva den kan antas å predikere. Det kan dermed være grunn til å forvente at Bus Story ikke kan predikere utviklingen av ordforråd og grammatiske ferdigheter spesielt godt. Likevel er det grunn til å undersøke dette særskilt. Det avgjørende for Bus Story-testens prediktive validitet er hvordan barnas språklige utvikling har vært i det året som har gått. Vi vet for eksempel at en god del fireåringer gjorde det bra på Bus Story, men svakt på testene for ordforråd og grammatikk. Hvis de samme barna skårer bra på ordforråd og grammatikk som femåringer, vil Bus Story likevel ha en viss prediktiv verdi. En mulig tolkning av en slik eventualitet kunne være at Bus Story tappet ferdigheter og egenskaper med betydning for den videre språkutvikling, men som på det tidspunktet likevel ikke manifesterte seg i høy skåre på enkelttester av ekspressivt og reseptivt språk. Tabell 4-5 viser sammenhenger mellom barnas resultater på Bus Story da de var fire år, og resultater på ulike språktester ett år senere.



Tabell 4-5: Korrelasjoner mellom Bus Story og tester for ordforråd og grammatikk administrert ett år senere.

Tester	Bilde- benevning	Grammatic Closure	BPVS	Trog 2-items	Ordf. WPSSI
Bus Story	.336**	.255**	.276**	.244**	.210**

\*\*, Korrelasjonen er signifikant på 0.01-nivå.

Undersøkelsen viser signifikante, positive korrelasjoner mellom resultater på Bus Story i fireårealderen og resultater på alle de fem språktestene ett år senere. Sammenhengene er imidlertid svake. Den sterkeste sammenhengen finner vi mellom Bus Story og Bildebenevning, testen for ekspressivt vokabular. Med en r-verdi på .336,  $p < .01$ , er det selv her tale om en svak sammenheng. Den svakeste sammenhengen finner vi mellom Bus Story og testen for dybdevokabular, en ekspressiv test, hvor  $r = .210$ ,  $p < .01$ . Det ser med andre ord ut til at fireåringenes resultater på Bus Story kun i beskjeden grad predikerer hvordan de gjør det på tester av ordforråd og grammatikk ved fem års alder.

#### 4.8 Kjennetegn ved gruppen av barn som skåret 0 poeng på Bus Story

Den empiriske analysen viser at så mange som ti barn ikke klarte å oppnå noen poeng overhodet på Bus Story. Ut fra et spesialpedagogisk ståsted er det naturlig å rette særskilt oppmerksomhet mot denne gruppen, og spørre hva den meget svake skåren kan skyldes. I denne sammenhengen er det spesielt interessant å se på disse barnas språkferdigheter for øvrig, slik de lar seg måle ved hjelp av andre tester i testbatteriet. Tabell 4-6 viser disse barnas testresultater sammenlignet med hele utvalget. Fra venstre til høyre vises hele utvalgets gjennomsnittsskåre på de ulike testene, 0-

gruppens gjennomsnittsskåre, 0-gruppens individuelle skårer, og til sist 0-gruppens avstand fra gjennomsnittet i standardavvik.

*Tabell 4-6: Gruppen som skåret null poeng på Bus Story Informasjon (0-gruppen) sammenlignet med hele utvalget mhp resultater på språktester, setningsminne og terningmønster (alder fire år)*

Test	Gjennomsnittsskårer hele utvalget (n=182)	Gjennomsnittsskårer 0-gruppen (n=10)	0-gruppens skårer individuelt* (n=10)	0-gruppens avstand fra utvalgets gjennomsnitt, i standardavvik
BPVS	42	35	22, 41, 45, 35, 17, 37, 28, 26, 42, 56	7/11 =0.65
Grammatic Closure	11	8.5	4, 11, 13, 13, 2, 10, 5, 6, 9, 13	2.5/4=0.60
Trog-2 items	37	29	14, 46, 30, 25, 8, 33, 34, 10, 30, 62	8/17=0.40
Bildebenevning	17	14.5	15, 18, 12, 17, 14, 13, 14, 8, 15, 21	2.5/3=0.80
Setningsminne	6.5	6	6, 9, 7, 5, 5, 8, 7, 3, 6, 5	0.5/2=0.25
Terningmønster	22	22.5	26, 22, 20, 20, 20, 23, 22, 20, 26, 26	0.5/3=0.15 (over snittet)

\*De individuelle skårene angis i samme rekkefølge for hver test. Barnet som oppnådde 22 på BPVS, fikk således 4 på Grammatic Closure, 14 på Trog 2-items og 15 på Bildebenevning, osv.

Av Tabell 4-6 framgår det at barna som skåret null poeng på Bus Story, som gruppe betraktet oppnår nokså normale resultater på testene for ordforråd og grammatikk, så vel som på testene for hukommelse og generelt evnenivå. De skårer riktignok lavt på alle språktestene, men ikke i noe tilfelle er gruppens gjennomsnitt så mye som ett

---

standardavvik under utvalgets gjennomsnitt. På Setningsminne skårer de et kvart standardavvik under gjennomsnittet, mens de på Terningmønster ligger litt over.

Ser vi på de individuelle verdiene til barna i denne gruppen, er bildet ganske uensartet. På hver av testene skårer noen av barna langt under gjennomsnittet, mens andre skårer likt eller over. På Bildebenevning, for eksempel, varierer skårene fra 8 til 21 poeng. Gjennomsnittet var her 17 poeng og standardavviket 3. Tre av de ti barna skårer mer enn ett standardavvik lavere enn gjennomsnittet på denne testen, mens tre skårer likt med eller over gjennomsnittet. Det er verdt å merke seg at ett av barna som fikk null poeng på Bus Story, skåret betydelig over hele utvalgets gjennomsnitt på alle de andre språktestene. Kun to av barna med null poeng på Bus Story gjorde det gjennomgående dårlig på alle de andre språktestene, dvs. at de skåret minst ett standardavvik lavere enn gjennomsnittet på hver av testene.

Resultatene viser altså at fireåringene som skåret null poeng på Bus Story, på langt nær gjorde det like dårlig på de andre språktestene. Det er også av interesse å undersøke hvordan språkutviklingen har vært for disse barna det påfølgende året. Tabell 4-7 viser hvordan barna som skåret null poeng på Bus Story, gjør det på språktester ett år senere, sammenlignet med gjennomsnittsverdier for hele utvalget. Et interessant spørsmål er om avstanden til gjennomsnittet er blitt større eller mindre i løpet av året som er gått.

*Tabell 4-7: Gruppen som skåret 0 poeng på Bus Story ved fire års alder (0-gruppen) sammenlignet med hele utvalget mhp resultater på språktester ett år senere*

Test	Gjennomsnittsskårer hele utvalget (n=172)	Gjennomsnittsskårer 0-gruppen (n=8)	0-gruppens skårer (n=8)	0-gruppens avstand fra utvalgets gjennomsnitt, i standardavvik
BPVS, t2	56	52.50	48, 52, 46, 50, 40, 57, 68, 57	3.5/11.5=0.30
Grammatic Closure, t2	14	14	14, 15, 13, 18, 17, 10, 15, 11	0/4=0
Trog-2 items, t2	50	46	46, 60, 52, 29, 43, 25, 44, 68	4/14.5=0.30
Bildebenevning, t2	21	18.50	20, 18, 18, 20, 19, 16, 14, 23	2.5/3.5=0.70

Resultatene fra testingen ett år senere viser at 0-gruppens avstand til utvalgsgjennomsnittet er blitt mindre på alle språktestene. På Grammatic Closure er avstanden helt utlignet. Det er verdt å merke seg at ett av de to barna som fikk null poeng på Bus Story, og som i tillegg lå mer enn ett standardavvik under utvalgets gjennomsnitt på alle de andre språktestene, har hatt en positiv utvikling fram til andre testtidspunkt. Dette barnet skårer fortsatt mer enn ett standardavvik under snittet på to av testene, men nøyaktig ett standardavvik under snittet på Grammatic Closure, og faktisk over snittet på BPVS. For det andre barnet som hadde en slik profil, foreligger ikke data fra andre testtidspunkt.

## 5. Oppsummering av funn, diskusjon og avslutning

### 5.1 Oppsummering av funn

#### **Problemstilling 1: Er skårene på Bus Story gode indikatorer på språklige ferdigheter hos norske fireåringer?**

Den empiriske undersøkelsen viser svake sammenhenger mellom barnas resultater på tester som måler ordforråd og grammatikk, og deres prestasjoner på Bus Story-testen. Resultatene gir ikke grunnlag for å anta at fireåringer som oppnår høye skårer på Bus Story, nødvendigvis også har et godt ordforråd og gode grammatiske ferdigheter. Motsatt gjelder det også at det ikke er grunnlag for å anta at fireåringer som gjør det svakt på Bus Story-testen, har et lite ordforråd og svake grammatiske ferdigheter. Den multiple regresjonsanalysen viser at ordforråd og grammatikk kun forklarer en liten del av variansen i barnas resultater på Bus Story. Dersom testen for ekspressivt vokabular innføres først, forklarer den cirka 14% av variansen. De tre øvrige språkvariablene bidrar da ikke med ekstra forklart varians. Forskjellene i barnas prestasjoner på Bus Story-testen kan med andre ord bare i liten grad forklares med forskjeller i ordforråd og grammatiske ferdigheter. Hele 86% av variansen skyldes andre faktorer. En nærmere undersøkelse av ti barn som fikk null poeng på Bus Story, viste at de som gruppe betraktet oppnådde svake, men likevel upåfallende resultater på de andre språktestene. Kun to av disse barna skåret gjennomgående svakt på alle språktestene. På dette grunnlaget synes det rimelig å konkludere med at resultatet på Bus Story ikke gir en god indikasjon på språklige ferdigheter hos norske fireåringer. Undersøkelsen viste også svake sammenhenger mellom Bus Story og tester for generelt evnenivå og umiddelbar hukommelse for talt språk. Resultater på testene Setningsminne og Terningmønster bidrar i liten eller ingen grad til å forklare variansen i resultater på Bus Story, utover det som de språklige testene kan forklare.

---

**Problemstilling 2: Hvor godt kan resultater på Bus Story predikere språklig utvikling for norske fireåringer?**

Den empiriske undersøkelsen viser signifikante, men svake sammenhenger mellom fireåringenes prestasjoner på Bus Story og deres resultater på tester for ordforråd og grammatikk når de er fem. Det er kun en svak tendens til at høy skåre på Bus Story ved fire års alder går sammen med et godt ordforråd og gode grammatiske ferdigheter ett år senere, og til at lav skåre på Bus Story i fire års alder går sammen med svakt ordforråd og svake grammatiske ferdigheter ett år senere. På bakgrunn av disse funnene synes det rimelig å konkludere med at fireåringers resultater på Bus Story i liten grad predikerer deres språkferdigheter ett år senere.

Testresultatene viser også at den gruppen fireåringer som gjorde det aller svakest på Bus Story, ikke har hatt en negativ språklig utvikling det påfølgende året, men tvert imot blitt mer lik gjennomsnittet. I et klinisk, forebyggende perspektiv gir ikke undersøkelsen grunn til å anse svake resultater på Bus Story i fire års alder som en særskilt risikomarkør. De fireåringene som mestret Bus Story aller dårligst, har fram mot fem års alder hatt en språkutvikling som gjør at de nærmer seg gjennomsnittlige resultater på tester av ordforråd og grammatiske ferdigheter.

## 5.2 Diskusjon

På grunnlag av de gjennomførte undersøkelsene synes det rimelig å konkludere med at det er vanskelig å si hva Bus Story ”tapper” hos norske fireåringer. En studie som denne kan naturligvis ikke gi noe endelig svar, men resultatene slik de foreligger, tilsier at Bus Story er en usikker test med tanke på å kartlegge norske fireåringers språk. Det presiseres igjen at det her er tale om hva *skårene* gir grunnlag for å si om barnas språkferdigheter, og ikke en generell vurdering av testens brukbarhet.

At språklige variabler forklarer så lite, må betegnes som overraskende. Hvorfor synes ordforråd og grammatiske ferdigheter å spille så liten rolle for hvordan fireåringene gjør det på Bus Story? En type forklaring kan knytte an til det som kalles lav konvergerende validitet. Testen fanger rett og slett ikke opp vesentlige sider ved barnas språkferdigheter. Slik testen administreres og skåres, gir ikke stort ordforråd eller grammatiske ferdigheter noe vesentlig fortrinn. Barna kan oppnå god uttelling ved å si i et enkelt språk hva de ser på bildene, slik det ble påpekt i denne oppgavens teoridel. De empiriske resultatene i undersøkelsen lar seg forene med en slik antagelse.

En annen type forklaring er ikke et alternativ, men kan være supplerende, og knytter an til det som kalles divergerende validitet. Det innebærer at andre faktorer som ikke er direkte relevante for det området man er interessert i, har stor innvirkning på testresultatet. Hvilke faktorer kan dette være? Hvis testen i liten grad måler ordforråd og grammatiske ferdigheter, hva måler den da? Spørsmålet kan også stilles som følger: Hva kjennetegner fireåringene som har gjort det bra på Bus Story? I det følgende skal to muligheter kort drøftes. Det blir da tale om teoretiske overveielser, som ikke kan belegges empirisk. En mulighet er at de barna som har gjort det best, er de som er mest fortrolig med sjangeren, de som er mest vant til å uttrykke seg i en narrativ modus. På et generelt nivå kan en slik hypotese forankres i et sosial-pragmatisk syn på språkutvikling (Tomasello, 2003). Det vektlegges her at språktilegnelse først og fremst skjer gjennom sosial interaksjon. Både formale og funksjonelle aspekter ved språket tilegnes gradvis ved at barn bruker språket og får tilbakemelding på sin språkbruk fra ulike samhandlingspartnere. Denne tilbakemeldingen bruker barn til å trekke slutninger og bygge opp kunnskap om språklige uttrykks kommunikative funksjon. I vår sammenheng kunne det knyttes an til litteratur som diskuterer betydningen av ulike former for språklig stimulering i hjemmemiljøet, og da spesielt bordsamtalens betydning (Snow & Aukrust, 1998; Snow & Beals, 2006). Snow og Beals (1998) mener at bordsamtalen kan være en spesielt viktig språklig sosialiserings- og læringsarena. En av grunnene er nettopp den muligheten til narrativ diskurs som denne settingen åpner opp for. Den narrativt

pregede bordsamtalen kan dreie seg om hverdagslige temaer, men kan like fullt være en relativt krevende og stimulerende språklig oppgave. Den som forteller må holde oversikt over aktører, settinger og hendelser, og spørsmål om motiver, årsaker og konsekvenser vil ofte være aktuelle. Bordsamtalen gir anledning til å utvikle ferdigheter som kreves for å mestre narrativ produksjon. For å vende tilbake til vårt materiale og spørsmål, kunne man spørre om det er slik at fireåringenes prestasjoner på Bus Story gjenspeiler deres spesifikke språklige stimulering, nærmere bestemt hvor vant de er hjemmefra til å bruke språket narrativt. Det kan tenkes at spesielt bordsamtaler utgjør en slags øvingsarena for den type oppgave som Bus Story stiller barna overfor, og at noen barn har fått mer øving enn andre. I denne sammenhengen er det verdt å merke seg at flere undersøkelser, også med data fra Norge, tyder på at det er store forskjeller mellom familier med hensyn til bordsamtalenes art og omfang (Snow og Aukrust, 1998; Snow og Beals, 2006). Det kan tyde på at noen barn rett og slett er mer vant til å fortelle, og dette kan tenkes å slå ut på resultater ved en test som Bus Story. I utgangspunktet er det rimelig å anta at de fleste norske fireåringer har begrenset erfaring med narrativ produksjon. Da kan forskjeller i omfanget av relevant erfaring fra hjemlig språkmiljø bety en god del for prestasjonene på en slik test. Noen av barna opplever kanskje Bus Story som en veldig uvant og fremmedartet oppgave, mens andre ”drar kjensel” på situasjonen og forstår raskt hva det går ut på.

Ut fra ovenstående resonnement kunne det være mulig å tenke seg at fireåringers skårer på Bus Story først og fremst sier noe om hvor vant de er til å fortelle, og lite om underliggende språkferdigheter som sådan. Det er imidlertid et problem knyttet til å bruke en slik forklaringsmodell på funnene i denne undersøkelsen. Forskningen tyder på at den type språklig stimulering som øvelse i narrativ produksjon gir, er generelt gunstig for barns språkutvikling (Snow & Beals, 2006.) Narrativ diskurs lærer barn hva det vil si å fortelle, og det stimulerer samtidig utviklingen av ordforråd og grammatiske ferdigheter (Hagtvet, 2004; Snow & Beals, 2006.) Det er med andre ord vanskelig å tenke seg en språklig stimulering som helt spesifikt gjør barna flinke til å fortelle, men som ikke samtidig utvikler språklige kjerneferdigheter som ordforråd og grammatikk. Hvis det er barn som har fått mye relevant



språkstimulering - det vil si øvelse i narrativ forståelse og produksjon - som har gjort det bra på Bus Story, burde vi sett en sterkere korrelasjon med de øvrige språktestene. Likevel bør det ikke helt ses bort fra at barn som er vant med sjangeren har hatt et visst fortrinn ved testingen med Bus Story, og at dette til en viss grad er uavhengig av språkferdigheter for øvrig. Det er mulig at slike barn har vært bedre i stand til å forstå hva oppgaven går ut på og hvordan den skal løses. Om de da ikke i utgangspunktet har bedre ordforråd og grammatiske ferdigheter, så kan det tenkes at de er flinkere til å bruke sine språkferdigheter til å løse den type oppgaver som Bus Story er et eksempel på. I så fall kan ”fortrolighet med narrativ språkbruk” være en variabel som forklarer en unik del av variansen i fireåringenes testresultater, om enn ikke en stor del.

En annen forklaring på funnene kan være at det er de mest utadvendte, ”pratsomme” og usjenerte fireåringene som har gjort det best på Bus Story. I så fall vil det si at slike personlighetstrekk er viktigere enn språkferdigheter per se for resultatene. Bus Story er en krevende oppgave sammenlignet med andre språktester, i den forstand at barnet får mye ansvar for fremdriften. Det må våge å eksponere seg og være språklig aktiv i en uvant situasjon, ofte overfor en voksen som han eller hun ikke kjenner godt fra før. Til sammenligning krever BPVS, for eksempel, kun at barna peker på riktig bilde, og Bildebenevning at de sier enkeltord. Karlsen (2008) antyder i sin masteroppgave om Bus Story at det muligens var uheldig at Bus Story ble administrert tidlig på første testdag, etter to reseptive tester som stilte relativt små krav til aktivitet fra barnets side. Det kan ha forsterket en tendens som nok vil gjøre seg gjeldende i mange tilfeller når man tester så små barn, nemlig at sjenanse spiller inn, og personlighetstrekk som utadvendthet/innadvendthet kan få svært mye å si for prestasjonen. Karlsen (2008) fant riktignok kun meget svake sammenhenger mellom ”emosjonalitet i testsituasjonen” og Bus Story-resultater. Hvert barn får en skåre for ”emosjonalitet i testsituasjonen” basert på testleders observasjon av atferd som kan indikere usikkerhet, ubehag og uro, som for eksempel neglebiting, motorisk uro, nøling, kremting, uttrykk for ”vondter” m.m. Men personlighetstrekk som gjør barnet lite tilbøyelig til å ta ordet, være pratsom og vise initiativ i en testsituasjon,

gjenspeiles ikke nødvendigvis i denne type atferd. Barn kan prestere svakt på Bus Story på grunn av sjenanse og innadvendthet, uten at det viser atferd som tydelig kan oppfattes og registreres som nervøsitet, usikkerhet og uro.

På grunnlag av denne oppgavens drøftinger og empiriske funn, kan det være grunn til å spørre om Bus Story er en egnet test ved kartlegging av språkferdigheter til så unge barn som det her er tale om. I teoridelen ble dette temaet berørt, og følgende spørsmål ble reist: Kan det tenkes at en narrativ test kan gi en god generell indikasjon på språkferdigheter bare under den forutsetning at oppgaven mestres rimelig bra, og at dette ikke er tilfelle allerede ved fire års alder? Er det mulig at oppgaven da forstås og mestres såpass dårlig at det blir svært mange tilfeldige og ukjente faktorer som virker inn, slik at det blir vanskelig å tolke en kvantitativ skåre? Resultatene i denne undersøkelsen tyder på at dette *kan* være tilfellet. Spørsmålet bør også vurderes i lys av den kunnskapen som forskningen har frambrakt om utviklingen av narrative ferdigheter hos barn. Hva kan egentlig forventes av fireåringer? Hovedtrekkene i denne forskningen blir derfor kort gjengitt i det følgende.

Utviklingen av barns narrative ferdigheter har vært gjenstand for omfattende forskning (Norbury & Bishop, 2003). Det mest omfattende forskningsprosjektet ble ledet av Slobin og Berman (1994). Her beskrives og sammenlignes karakteristiske trekk ved narrativ produksjon hos henholdsvis tre, fem, syv og ni år gamle barn i ulike land. Forskerne baserte seg på data samlet inn ved hjelp av testen "Frog, Where are you?". Her skal barna fortelle en historie ut fra 24 bilder, det er altså ikke en gjenfortellingsoppgave. Et viktig funn hos Slobin og Berman (1994) var en felles utviklingsprofil på tvers av land og språk med hensyn til narrative ferdigheter. "Ferdigheter" ble i denne studien knyttet til barnas behandling av tre essensielle narrative strukturelementer: Utgangspunkt/setting med presentasjon av aktører og problem eller mål, forsøk på å løse problemet eller nå målet, og det endelige utfallet. Forskerne fant at treåringene fortalte episodisk og springende. Deres ytringer var i svært liten grad styrt av hensynet til kausal og temporal logikk i framstillingen. Mer

enn 80% av treåringene hadde ingen av de tre strukturelementene med i sine fortellinger. Slobin og Berman (1994) vurderte at kun en av 60 treåringer produserte det som kunne sies å være en narrativ. Det var en klar tendens til at barna beskrev bildene enkeltvis, og det var store variasjoner når det gjaldt hva de valgte å snakke om. Typisk for treåringene var også at de opptrådte mer personlig og interaktivt i testsituasjonen enn eldre barn. De var tilbøyelige til å vandre fra historiens innhold og fortelle om personlige opplevelser med vage forbindelser til det som historien handlet om, og til å påkalle testlederens oppmerksomhet i form av utrop som "se!" og "se på den!". Det er meget interessant at forskerne i Slobin og Bermans prosjekt hadde store problemer med selve datainnsamlingen blant treåringene. Det rapporteres at *"large numbers of 3-year-olds were interviewed who were not included in our database, since they either refused or failed to talk about the picture book at all"* (Slobin & Berman, 1994, s. 58-60). Slobin og Berman (1994) viser også til en upublisert doktoravhandling (Renner, T., 1988, ref. i Slobin & Berman, 1994), hvor forskeren etter 60 testøkter med treåringer satt igjen med tolv narrativer som kunne brukes til forskning. De øvrige var for korte og/eller uforståelige til at de kunne transkriberes. Det synes med andre ord å være store problemer forbundet med å bruke narrative tester på barn så unge som tre år. En stor andel av barna på dette alderstrinnet klarer ikke å forholde seg til oppgaven. Hva så med fireåringene? På bakgrunn av de erfaringene som nettopp er gjengitt, kan det virke overraskende at en del forskere overhodet ikke nevner problemer med å få fireåringer med ekspressive språkvansker til å gjennomføre en narrativ test (Bishop & Edmundson, 1987; Paul & Smith, 1993). Slobin og Bermans undersøkelse inkluderte også et utvalg fireåringer. Bildet her var nokså uensartet. Noen liknet mest på treåringene, ved at de beskrev bilde for bilde i et svært enkelt språk. Andre ga mer helhetlige, tematisk motiverte framstillinger med bedre utviklet syntaks og et rikere ordforråd. De befant seg med andre ord – ikke overraskende – i en mellomposisjon mellom tre- og femåringene, både mht. strukturbevissthet og språklige ferdigheter. Det bør da bemerkes at heller ikke femåringene - som gruppe betraktet - i noen særlig grad viste evne til å binde individuelle ytringer sammen til en helhet, og klargjøre kausale og temporale forhold.

Bus Story er nok en enklere test enn "Frog, where are you?", blant annet fordi barna ved Bus Story får høre historien først. Like fullt styrker etter min oppfatning den forskningen som det her vises til en "mistanke" om at det er i tidligste laget å bruke narrative tester, kvantitativt skåret, i kartleggingen av fireåringers språkferdigheter. I denne forbindelse er verdt å merke seg at såpass mange av barna i vårt utvalg skåret null poeng på Bus Story, og at dette var barn som stort sett lå innenfor normalområdet på øvrige språktester.

### 5.3 Avslutning: Implikasjoner og fremtidige forskningsspørsmål

Sammenhengene mellom fireåringers resultater på Bus Story og deres språkferdigheter for øvrig ble i denne undersøkelsen funnet å være svake. Barnas prestasjoner synes i liten grad å gjenspeile deres ordforråd og grammatiske ferdigheter. Heller ikke kognitive ferdigheter som generelt evnenivå og hukommelse for talt språk, slik de har latt seg operasjonalisere i denne oppgaven, synes å være av noen særlig betydning for hvordan oppgaven mestres. Det er ikke uventet at flere faktorer spiller inn. Det er imidlertid overraskende at de undersøkte variablene - som omfatter kjerneelementer i språklig kompetanse - synes å bety så lite. Det gjenstår over 80% uforklart varians når antatt viktige språklige og kognitive variabler har fått forklare sitt. Funnene gir grunnlag for å hevde at det er vanskelig å tolke de skårene som fireåringene oppnår på Bus Story. Det er vanskelig, for å bruke Kanes (1992, s. 27) formulering: *"to explain the meaning of the score, and, thereby, to make at least some of the implications of the score clear"*. På dette grunnlaget er det nærliggende å stille spørsmål ved testens egnethet i kartleggingen av fireåringers språkferdigheter. Det kan tenkes at dersom narrative tester skal brukes i kartleggingen av fireåringers språk, er en kvalitativ tolkning av prestasjonen mest aktuell.

I diskusjonen av funnene ble det antydnet at den lave korrelasjonen mellom Bus Story og de øvrige språktestene kan skyldes at den narrative oppgaven ikke mestres godt

nok av fireåringene. Det kan tenkes at fireåringer ikke er modne nok til å forholde seg adekvat til oppgaven, og at resultatene da i stor grad blir influert av mer eller mindre tilfeldige og irrelevante faktorer. På dette grunnlaget kunne man formulere en hypotese om at prestasjoner på Bus Story vil være en bedre test å bruke når barna er litt eldre, for eksempel seks år. Flere forskere peker på at det synes å skje et kvalitativt sprang i utviklingen av narrative ferdigheter i femårsalderen (Karmiloff & Karmiloff-Smith, 2001). Det er mulig at Bus Story-resultater på et høyere alderstrinn enn fire år er en bedre indikator for generelle språkferdigheter, og slik sett er en mer interessant test å bruke i kartlegging. Dette kunne være tema for en fremtidig undersøkelse: Vil Bus Story-resultater vise en sterkere sammenheng med øvrige språkferdigheter, som ordforråd og grammatisk kompetanse, når det er litt eldre barn som testes, og selve oppgaveformen mestres bedre?

Innenfor rammen av Språkprosjektet er det også mulig å fortsette undersøkelsen av Bus Story-testens prediktive egenskaper. Et spørsmål kan være om Bus Story, til tross for de svake sammenhengene med språklige kjerneferdigheter som kommer fram i denne undersøkelsen, likevel "tapper" en evne eller ferdighet som har betydning for senere språklig og/eller faglig mestring, for eksempel tilegnelsen av lese- og skriveferdigheter. Dette kan undersøkes gjennom enkle korrelasjonsanalyser. Hvilken sammenheng er det, for eksempel, mellom fireåringers resultater på Bus Story og deres lese- og skriveferdigheter i syv års alder? Undersøkelser av slike spørsmål kan ses som ytterligere bidrag til å vurdere testens validitet. Ut fra funnene i denne oppgaven, vil det ikke være overraskende om sammenhengene mellom Bus Story-resultater i fire års alder og senere språklig og generell skolefaglig mestring er svake. Skulle man derimot finne at det er visse sammenhenger, får spørsmålet om hva Bus Story faktisk måler fornyet interesse.

## Kildeliste

- Aukrust, V.G. & Snow, C.E. (1998). Narratives and explanations during mealtime conversations in Norway and the US. *Language in Society*, 27, 221-246.
- Bachman, L.F., (2006). Building and Supporting a Case for Test Use. *Language Assessment Quarterly*, 2(1), 1-34.
- Befring, E: ”Kvantitativ metode” (Sist oppdatert: 24. mai 2009). De nasjonale forskningsetiske komiteer. [Online]. Tilgjengelig på <http://etikkom.no/no/FBIB/Introduksjon/Metoder-og-tilnarminger/Kvantitativ-metode/>. Lesedato: 14.06.09.
- Befring, E. (2007). *Forskningsmetode med etikk og statistikk*. Oslo: Det Norske Samlaget.
- Berman, R.A., & Slobin, D.I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale New Jersey: Lawrence Erlbaum Associates.
- Bishop, D.V.M., (2003)/Lyster, S.A.H., & Horn, E. (2009). *Norsk håndbok for Test for Reception of Grammar. Trog-2 manual*. Gøteborg: Harcourt Assessment.
- Bishop, D.V.M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry*, 31, 1027-1050.
- Bishop, D.V.M., & Edmundson, A. (1987). Language impaired 4-year olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, Vol. 52, 156-173.
- Bloom, L. & Lahey, M., (1978). *Language Development and Language Disorders*.
- Botting, N., (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, 18,1.
- Brown, R. (1973). *A First Language. The Early Stages*. Cambridge, Massachusetts: Harvard University Press.

Child Language & Learning prosjektsøknad (2007). *The Nature and development of language and communication skills in pre-school children*. Søknad om midler til frittstående prosjekter – samfunnsvitenskap (FRISAM), Norges Forskningsråd.

Cohen, J. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, N.J.: Lawrence Erlbaum.

Dunn, L.M., Dunn, L.M., Whetton, C. & Burley, J. (1997). *The British Picture Vocabulary Scale*, Second edition, nferNelson, London.

Ellis-Weismer, S. E. (1997). Intervention for children with developmental language delay. I Bishop, D.V.M. & Leonard, L. B. (red) *Speech and language impairments in children. Causes, Characteristics, Intervention and Outcome*. New York: Psychology Press.

Enger, H-O., & Kristoffersen, E.K. (2000). Innføring i norsk grammatikk. Morfologi og syntaks. Oslo: Cappelen Akademisk Forlag.

Feagans, L. & Appelbaum, M.I. (1986). Validation of Language Subtypes in Learning Disabled Children. *Journal of Educational Psychology*, Vol 78., No 5, 358-364.

Fey, M.E., Catts, H.W., Proctor-Williams K., Tomblin J.B. & Zhang, X. (2004). Oral and Written Story Composition Skills of Children With Language Impairment. *Journal of Speech, Language, and Hearing Research*, Vol. 47.

Fiske, D.W. (2002). Validity for What? I Braun, H.I., Jackson, D.N. & Wiley, D.E. (red.), *The Role of Constructs in Psychological and Educational Measurement*. New Jersey & London: Lawrence Erlbaum Associates.

Hagtvet, B. E. (2004). *Språkstimulering. Tale og skrift i førskolealderen*. Oslo: Cappelen Akademisk Forlag.

Heiman, G.W. (1996). *Basic Statistics for the Behavioral Sciences*. 2<sup>nd</sup> edition. Boston: Houghton Mifflin Company.

- 
- Hitchcock D. & Verheij, B. (2006). Introduction. I Hitchcock D. & Verheij, B. (red.) *Arguing on the Toulmin model. New Essays in Argument Analysis and Evaluation*. Dordrecht: Springer.
- Howlin, P. & Kendall, L. (1991). Assessing children with language tests – which tests to use? *British Journal of Disorders of Communication*, 26.
- Haakanes, B. (2008). *Grammatiske ferdigheter hos norske fireåringer*. Masteroppgave, Universitetet i Oslo, Oslo.
- Karlsen, J. (2008). *Narrativ produksjon hos fireåringer*. Masteroppgave, Universitetet i Oslo, Oslo.
- Kane, M.T. (1992). An Argument-Based Approach to Validity. *Psychological Bulletin*, Vol. 112, No. 3, s. 527-535.
- Karmiloff, K. & Karmiloff-Smith, A. (2001). *Pathways to language. From fetus to adolescent*. Cambridge: Harvard University Press.
- Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes*, 1 (1), 61-85.
- Kirk, S.A., McCarthy, J.D. & Kirk, W.D. (1972). *Illinois Test of Psycholinguistic Abilities*. Revidert utgave. Håndbok med instruksjoner og normer. Norsk utgave ved Leonard, L.B. (1998). *Children with specific language impairment*. Cambridge, Massachusetts: MIT Press.
- Kristoffersen, E.K., Simonsen, H.G. & Sveen, A. (red) (2008). *Språk. En grunnbok*. Oslo: Universitetsforlaget.
- Gjessing, H.J., & Nygaard, H.D. m.fl. Skolepsykologi – materiellservice.
- Kleven, T.A. (2002a). Begrepsoperasjonalisering. I Lund, T. (red.) *Innføring i forskningsmetodologi*. Oslo: Unipub Forlag.
- Kleven, T.A. (2002b). Ikke-eksperimentelle design. I Lund, T. (red.) *Innføring i forskningsmetodologi*. Oslo: Unipub Forlag.



Kunnskapsdepartementet. (2006-2007)...*og ingen sto igjen. Tidlig innsats for livslang læring*. St.meld. nr 16. Oslo: Kunnskapsdepartementet.

Lewis-Beck, M.S. (1980). *Applied Regression. An Introduction*. London: Sage Publications.

Liles, B.Z. (1993). Narrative Discourse in Children With Language Disorders and Children With Normal Language: A Critical Review of the Literature. *Journal of Speech and Hearing Research*, Vol. s. 36, 868-882.

Lund, T. (2002). Generaliseringsproblematikk. I Lund, T. (red.) *Innføring i forskningsmetodologi*. Oslo: Unipub Forlag.

Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher*, Vol. 18, No. 2, s. 5-11.

Messick, S. (1980). Test Validity and the Ethics of Assessment. *American Psychologist*, Vol 35, No. 11, 1012-1027.

Miles, S. & Chapman S.C. (2002). Narrative Content as Described by Individuals With Down Syndrome and Typically Developing Children. *Journal of Speech, Language, and Hearing Research*, Vol. 45, s. 175-189.

Mislevy, R.J. (2003). Substance and Structure in assessment arguments. *Law, Probability and Risk*, 0, s. 1-22.

NESH publisering 2006, *Forskningsetiske retningslinjer for samfunnsvitenskap, jus og humaniora*, lesedato: 25.05.09.

<http://www.etikkom.no/retningslinjer/NESHretningslinjer>

Norbury, C.F. & Bishop, D.V.M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders*. Vol 38, No 3, 287-313.

Nordeide, I. (2008). *Reliabilitet ved bruk av Renfrew Bus Story Test*. Masteroppgave, Universitetet i Oslo, Oslo.

Ottem, E. & Frost, J. (2005). *Språk 6-16. Screening Test-manual*. Oslo: Bredtvet Kompetansesenter.

- 
- Pankratz, M.E., Plante, E., Vance, R. & Insalaco, D.M. (2007). The Diagnostic and Predictive Validity of the Renfrew Bus Story. *Language, Speech, and Hearing Services in Schools*, Vol. 38, s. 390-399.
- Paul, R. & Smith, R.L. (1993). Narrative Skills in 4-Year-Olds With Normal, Impaired, and Late-Developing Language. *Journal of Speech and Hearing Research*, Vol. 36, s. 592-598.
- Plante, E., & Vance, R. (1994). Selection of Preschool Language Tests: A Data-Based Approach. *Language, Speech, and Hearing Services in Schools*, Vol. 25, s. 15-24.
- Plunkett, K. & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, vol. 48, no. 1, p. 21-69.
- Renfrew, C. (1997). *Bus story test. A test of narrative speech*. (4. ed.) Bicester: Winslow press. Bearbeidet til norsk for forskningsformål ved Institutt for spesialpedagogikk, Universitetet i Oslo.
- Ringdal, K. (2007). *Enhet og mangfold*. Bergen: Fagbokforlaget.
- Shadish, W.R., Cook, T.D. & Campbell D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York-Boston: Houghton Mifflin Company.
- Simonsen, H.G. (2001). Past tense acquisition in Norwegian. I Simonsen, H.G. & Endresen, R.T. (red.) *A Cognitive Approach to the Verb. Morphological and Constructional Perspectives*. New York: Mouton de Gruyter.
- Sireci, S.G. (2007). On Validity Theory and Test Validation. *Educational Researcher*, Vol. 36, No. 8, s. 477-481.
- Sireci, S.G. (1998). The Construct of Content Validity. *Social Indicators Research* 45, s. 83-117.
- Snow, C.E. & Beals, D.E. (2006). Mealtime Talk That Supports Literacy Development. *New Directions for Child and Adolescent Development*, no. 111.

---

Statistisk Sentralbyrå (2008). *Statistisk årbok 2008*. Hentet 5. november 2008, fra <http://www.ssb.no/aarbok/emne05.html>

Stothard, S.E., Snowling, M.J., Bishop, D.V.M., Chipchase, B.B., & Kaplan C.A. (1998). Language-Impaired Preschoolers: A Follow-Up Into Adolescence. *Journal of Speech, Language, and Hearing Research*, Vol. 41, 2, s. 407-418.

Templin, M.C. (1957). *Certain Language Skills in Children. Their Development and Interrelationships*. Minneapolis: The University of Minnesota Press.

Tomasello, M. (2005). *Constructing a language: A usage-based approach to language acquisition*. Cambridge, MA: Harvard University Press.

Toulmin, S. (2003). *The Uses of Argument*. Cambridge: Cambridge University Press.

Wechsler, D. (2002). *WPPSI-III: Wechsler preschool and primary scale of intelligence*, 3<sup>rd</sup> edition. Manual. (Svensk versjon ved Eva Tidemann 2005 ed.) Psychological Corp., Harcourt Assessment.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence – Revised*. San Antonio, TX: The Psychological Corporation: Harcourt Assessment.  
Norsk versjon ved Rostad, A.M. & Hoven, G. (2002).

Williams, K.T. (2007). *Expressive Vocabulary Test*, 2<sup>nd</sup> ed., (EVT-2). Pearson Assessments. NCS Pearson inc.

---

**Vedlegg 1: Bus Story-testen (tekst og instruksjon)**

Det var en gang en veldig rampete buss. Mens sjåføren prøvde å reparere bussen, bestemte den seg for å rømme.

Den kjørte langs veien ved siden av et tog. De laget rare grimaser til hverandre og kjørte om kapp. Men bussen måtte fortsette alene, fordi toget kjørte inn i en tunnel. Bussen skyndte seg inn i byen hvor den møtte en politimann som blåste i fløyten sin og ropte: ”Stopp buss!”

Men den rampete bussen brydde seg ikke om det og kjørte videre ut på landet. Den sa: Jeg er lei av å kjøre på veien.” Så hoppet den over et gjerde. Den møtte en ku som sa: ”Mø, jeg tror ikke mine egne øyne”.

Bussen raste nedover bakken. Den forsøkte å stoppe, da den så at det var et vann i bunnen av bakken. Men den visste ikke hvordan den skulle sette på bremsene. Så falt den i vanndammen med et plask og satt fast i gjørma. Da sjåføren fant ut hvor bussen var, ringte han etter en heisekran som kunne dra bussen opp og få den tilbake på veien igjen.

**Instruksjon:** - Se i boka sammen med barnet - La barnet se på bildene litt før du forteller historien til bildene - Si: *”Nå skal jeg fortelle en historie om denne bussen (pause - pek på bildene i rekkefølge). Når jeg er ferdig med å fortelle, vil jeg at du skal fortelle denne historien om bussen”*. (fortell historien til bildene) - Sett på lydopptaker - Si: *”Nå kan du fortelle historien om bussen til meg. Det var en ...?”* - Snu sidene etter hvert som barnet forteller. Kan indirekte prompte barnet ved å si *”jaa...?”*, *”og så...?”* eller *”hva skjedde da...?”*, dersom det stopper opp i historien. Fortsett med å vise bildene uten å kommentere, selv om barnet ikke sier noe.

---

**Vedlegg 2: Transkripsjon av testrespons som oppnådde omtrent gjennomsnittlig skåre på informasjon (13 poeng)**

Testleder (T): *Det var en...*

Barn (B): Gang...en rampete buss

T: mm...

B: og han skulle kjøre

T: Ja...og hva skjedde da da?

B: mm...Han blei gla. Det blei rare grimaser

T:mm...akkurat sånn ja (*barnet gjør grimaser*). Og hva skjedde da?

B: Toget kjørte alene i en tunnel

T:mm...og så da?

B: Han blitt politimannen ropte stopp

T: ja...

B: Og de gjorde sånn. Han grønne gjorde sånn.

T:mm...

B:de skulle liksom stanse bussen

T:mm...og så hva skjedde her da?

B: ehm...han han kjørte der...hoppa'n over gjerdet også møtte han en ku

T:mm...

B: Og så fal...og så kjørte han ned bakken også f...plask så satt han fast i havet. Det var klissete oppi'n.

T:m...

B: Men vi ser ikke at det er klisset til. Også kom en ehm sånn ehm lei buss

T:m...

B: Og han stansa ehm der som han kjørte

T:jah...

B: Også bare plask

T: Plask sa det, ja

B: Hva skal vi gjøre nå?